AL/CF-TR-1995-0002

ARMSTRONG LABORATORY

PROTOTYPING AND INTERFACE/WORKLOAD EVALUATION OF A MULTISENSOR TARGET ACQUISITION SYSTEM (U)

Gilbert G. Kuperman

CREW SYSTEMS DIRECTORATE
HUMAN ENGINEERING DIVISION
WRIGHT-PATTERSON AFB OH 45433-7022

Alice D. Friedman

LOGICON TECHNICAL SERVICES, INC.
P. O. BOX 317258
DAYTON, OHIO 45431-7258

JULY 1994

INTERIM REPORT FOR THE PERIOD MAY 1992 TO JULY 1994

AIR FORCE MATERIEL COMMAND
WRIGHT-PATTERSON AIR FORCE BASE, OHIO 45433-6573

NOTICES

When US Government drawings, specifications, or other data are used for any purpose other than a definitely related Government procurement operation, the Government thereby incurs no responsibility nor any obligation whatsoever, and the fact that the Government may have formulated, furnished, or in any way supplied the said drawings, specifications, or other data, is not to be regarded by implication or otherwise, as in any manner licensing the holder or any other person or corporation, or conveying any rights or permission to manufacture, use, or sell any patented invention that may in any way be related thereto.

Please do not request copies of this report from the Armstrong Laboratory. Additional copies may be purchased from:

National Technical Information Service
5285 Port Royal Road
Springfield, Virginia 22161

Federal Government agencies and their contractors registered with the Defense Technical Information Center should direct requests for copies of this report to:

Defense Technical Information Center
Cameron Station
Alexandria, Virginia 22314

TECHNICAL REVIEW AND APPROVAL

AL/CF-TR-1995-0002

This report has been reviewed by the Office of Public Affairs (PA) and is releasable to the National Technical Information Service (NTIS). At NTIS, it will be available to the general public, including foreign nations.

This technical report has been reviewed and is approved for publication.

FOR THE COMMANDER

KENNETH R. BOFF, Chief
Human Engineering Division
Armstrong Laboratory

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. AGENCY USE ONLY *(Leave blank)*

2. REPORT DATE
July 1994

3. REPORT TYPE AND DATES COVERED
Interim Report May 92 - July 94

4. TITLE AND SUBTITLE
Prototyping and Interface/Workload Evaluation of a Multisensor Target Acquisition System

5. FUNDING NUMBERS
C: F41624-94-D-6000
PE: 62202F
PR: 7184
TA: 10
WU: 44

6. AUTHOR(S)
Gilbert G. Kuperman
Alice D. Friedman

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
Logicon Technical Services, Inc.
P.O. Box 317258
Dayton, OH 45431-7258

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)
Armstrong Laboratory, Crew Systems Directorate
Human Engineering Division
Human Systems Center
Air Force Materiel Command
Wright-Patterson AFB, OH 45433-7022

10. SPONSORING/MONITORING AGENCY REPORT NUMBER
AL/CF-TR-1995-0002

11. SUPPLEMENTARY NOTES

12a. DISTRIBUTION/AVAILABILITY STATEMENT
Approved for public release; distribution is unlimited

12b. DISTRIBUTION CODE

13. ABSTRACT *(Maximum 200 words)*

A prototype interface for multisensor target acquisition, synthetic aperture radar (SAR), and forward looking infrared sensors (FLIR) was evaluated by fourteen rated USAF subject matter experts. Four levels of multisensor integration (low and medium resolution SAR with manual target designation, the addition of a high resolution SAR mode, the addition of an automatic target cuer, and the addition of a FLIR sensor) were explored. Projected workload was assessed using the Subjective Workload Assessment Technique (SWAT) and the operator interface was evaluated using a structured rating scale tool.

14. SUBJECT TERMS
Synthetic aperture radar, SAR, forward looking infrared, FLIR, workload, SWAT, operator interface

15. NUMBER OF PAGES
113

16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
UNCLASSIFIED	UNCLASSIFIED	UNCLASSIFIED	UNLIMITED

NSN 7540-01-280-5500

Standard Form 298 (Rev. 2-89)
Prescribed by ANSI Std. Z39-18
298-102

This page intentionally left blank.

PREFACE

This effort was conducted under exploratory development Work Unit 7184 10 44, "Advanced Strategic Cockpit Engineering and Research." It was performed by the Crew Systems Integration Branch, Human Engineering Division, of the United States Air Force Armstrong Laboratory (AL/CFHI), Wright-Patterson Air Force Base, Ohio, and supported by Logicon Technical Services, Inc. (LTSI), Dayton, Ohio, under Contract F416-94-D-6000. Mr. Robert Linhart was the Contract Monitor and Mr. Gilbert Kuperman was the Work Unit Manager.

This activity was conceived as a predecessor to advanced development efforts to be conducted in support of the Air Force Theater Missile Defense Attack Operations Program. Lt Col Mike Tankersley (ASC/LAAT) is the Program Manager. Maj Wayne Miller (HQ ACC/DRA) is the Attack Operations Integrated Product Team Leader.

Special thanks are due to Mr. Kenneth Crum (LTSI) who provided the software support for the simulation and to Mrs. Iris Davis (LTSI) who conducted the concept demonstrations and Subject Matter Expert (SME) interviews. The professionalism and technical expertise of the fourteen Air Force officers who served as SMEs during the conduct of the study merits special acknowledgement.

TABLE OF CONTENTS

SECTION	TITLE	PAGE
	LIST OF FIGURES	vi
	LIST OF TABLES	vii
1	INTRODUCTION	1
2	CONCURRENT ENGINEERING AND THE RAPID PROTOTYPING PROCESS	3
3	RAPID PROTOTYPING	7
4	INTERFACE DESIGN	10
5	SENSOR FUSION AND INTEGRATION	16
6	SIMULATION	19
7	METHODS	21
	FACILITY	21
	SUBJECTS	23

	OVERVIEW	23
	MISSION SCENARIOS	24
	AVIONICS ALTERNATIVES	25
8	ANALYSIS AND RESULTS	30
	WORKLOAD	30
	OPERATOR INTERFACE EVALUATION	35
	AVIONICS CONCEPT EVALUATIONS	36

TABLE OF CONTENTS (Cont.)

SECTION	TITLE	PAGE
9	CONCLUSIONS	47
	REFERENCES	49
	GLOSSARY	54
	APPENDIX A: WORKLOAD	56
	APPENDIX B: RATING SCALE SME COMMENTS	73

LIST OF FIGURES

FIGURE	TITLE	PAGE
1	Building Block Approach to Sensor Fusion	16
2	SABER Crew Station	21
3	Family of Paired t-Tests Inter-task Comparisons	33
4	Mean SME Rating for General Areas of Consideration	35
5	Mean SME Ratings for Detection	38
6	Mean SME Ratings for Location	39
7	Mean SME Ratings for Identification	39
8	Mean SME Ratings for Workload	41
9	Mean SME Rating for Training	42
10	Mean SME Ratings for Crew Confidence	42
11	Mean SME Ratings for Situational Awareness	43

12	Mean SME Ratings for Mission Pacing	43
13	Mean SME Ratings for False Alarms	44
14	Mean SME Ratings for the Sensor Management Subsystem	45
15	Mean SME Ratings for Man-in-the-Loop Involvement	46

LIST OF TABLES

TABLE	TITLE	PAGE
1	Mission Run: Avionics Alternatives	25
2	SWAT Values	31
3	SWAT ANOVA Table	32
4	Results of Family t-Test for Identification	40
5	Results of Family t-Test for Workload	41
6	Results of Family t-Test for False Alarms	44
7	Results of Family t-Test for "Man-in-the-loop"	46

THIS PAGE INTENTIONALLY LEFT BLANK

SECTION 1

Introduction

Technology insertion programs present unique challenges to the crew system design team. New/maturing subsystem capabilities must be integrated with existing subsystems. The Air Force Theater Missile Defense Attack Operations (TMD AO) Program is exploring the enhancement and/or addition of target acquisition sensors and automatic target cueing and recognition (ATC/ATR) capabilities to the F-16C and F-15E fighter/attack weapon systems. The human factors engineering support to TMD AO is concerned with how best to integrate these new subsystem capabilities.

The integration problem is multi-faceted. It encompasses consideration of operator capabilities and limitations (i. e., crew workload) as well as system performance improvements. It also must address the controls and displays which compose the operator interface. These issues are elevated in importance when the task to be performed is that of target acquisition. Timelines are often highly compressed, enemy threat systems may be of great concern, and uncertainty is introduced by the demands of fratricide avoidance and minimization of collateral damage. These concerns are in addition to the perceived costs associated with missing targets or wasting weapons.

Past studies of target acquisition tasks that have evaluated the performance of the operator or that have focused on the performance capabilities expected to be improved upon or achieved, have actually evaluated the hardware (i. e., the ability of the sensor to produce an image that would improve performance due to improved sensitivity and resolution) rather than the workload involved in the use of the system. Yet, there is also a compelling need to evaluate the Human-System Interface (HSI) in order to assess the

changes on the HSI inherent with the adoption of the new technology, and make changes accordingly. Ideally, the earlier the HSI is evaluated in the system design process, the better. This is a key concept to the rapid prototyping techniques employed by concurrent engineering (CE) principles, as presented in this study.

Typically, testing of sensor fusion or integration concepts has focused on the performance of a designated avionics system or specific concept, and has not included an evaluation of the effects of progressive sensor integration on the HSI particularly with regard to workload. Using techniques for quantitative measurement of operator workload will allow selection of the technologies to ensure that final integration results in reduced workload and in improvement of the overall effectiveness or performance of the craft. (Vikmanis, 1987) This can be best achieved within the context of a CE environment and using rapid prototyping tools as demonstrated in this study.

The main objective of this project is to evaluate the effect of sensor integration on the HSI, specifically that of operator workload, and demonstrate the use of rapid prototyping of the HSI within the framework of a CE environment. While workload can be the determining factor in the performance of the operator, the HSI is a determining factor of workload. A progressive integration of sensors and modalities during a task loading process (mission scenario) allows selective evaluation of the HSI (concerning workload) for subsequent evaluation of iterations employed by rapid prototyping. It is on this premise that this study was undertaken.

SECTION 2

Concurrent Engineering and the Rapid Prototyping Process

Rapid prototyping is one of the more powerful tools of CE. This Section presents some current ideas and writings in the area of CE. Although the emphasis is on methods for prototyping the Human-Computer Interface (HCI), including a description of a structured approach to the iterative design of the HCI, it is applicable to any Human-System Interface (HSI). In addition, brief summaries of some other recent applications of rapid prototyping specific to crew station design are presented. (For the purpose of this section, HCI and HSI may be considered to be interchangeable.)

Mention must be made of the need to evaluate the HSI early in the development of new technologies in order to assess the changes inherent with the adoption of the new technology. After all, the HSI may be a determining factor in the overall performance capability of a system. For example, although the designer may produce a perfectly functioning sensor sub-system, operators may require that accommodation be made for their capabilities and limitations in the overall system. Currently, many of the cockpit designs do not allow for an optimal HSI. Furthermore, recent advances that have replaced old technology with newer, more advanced systems, have added to the problem of evaluating limitations inherent in changes to the HSI (due not only to the complexity of the systems and tasks involved, but also from the complexity inherent within operator workload and performance evaluation). These concepts are elucidated in a growing body of literature (Gopher & Donchin, 1986; Kalawsky, 1987; Kantowitz & Casper, 1988; Matthes, 1987; Reader, 1987; Sexton, 1988; Tsang & Vidulich, 1989; Walker, 1987). As will be discussed

later, the earlier the HSI is evaluated in the system design process, the more likely it is that it will adequately support all system functions.

Unfortunately, integration of new or additional avionics is often determined by newer, emerging technologies (e.g., "technology push") or by new or additional operational requirements (e.g., "requirements pull"). The engineering emphases are usually placed on the anticipated performance capabilities. The effect on the HSI, critical to the employment of the system, is often overlooked. This is where the concepts of rapid prototyping and CE become major considerations.

Concurrent engineering (also known as system engineering, simultaneous engineering, producibility engineering, or integrated product development) is defined by the Institute for Defense Analysis (Winner et al., 1988) to be a "systematic approach to the integrated, concurrent design of products and their related processes, including manufacture and support." Winner et al. (1988) stress, however, that concurrent engineering is not just concerned with product design, but includes the added requirement that the design is developed using a cost-effective engineering process.

More recently, Shumaker (1990) from the Concurrent Engineering Office, Manufacturing Technology Directorate (CEO, ManTech), defined CE as "the integrated design of the product, manufacturing, and support processes together with emphasis on efficiency, improved quality, and reduced cost," and noted the importance of CE as a means to increase combat capability. With regard to the USAF, new and changing threats mean the necessity to "respond more quickly in the development of new weapons systems required to counter those threats and, thereby help to ensure our nation's defense" (Houpt, 1990). CE relies on a range of methods to improve the acquisition process with regard to time, as well

as cost. (Within the acquisition process, CE begins with Milestone 0/Program Initiation and continues throughout the life cycle of the system.)

Shumaker (1990) noted that the cost of systems is increasing as the complexity increases, with the cost of airborne electronics accounting for up to 50% of recent aircraft cost. In fact, there is an exponential increase in costs when making changes as the system "matures," such that changes made earlier are less costly. With the increased complexity of newer systems, some of these changes are related to compatibility between the system and the operator.

Taking this into consideration, lower acquisition and life cycle costs are part of the ManTech's goals (Houpt, 1990) to be achieved by the mid to late '90s, and include:

1. Reduce concept to production time by 30%
2. Reduce the number of unplanned engineering changes by 50%
3. Reduce the cost of reliability and maintainability enhancements by 50%
4. Reduce sustaining engineering by 50%.

To achieve the improvements desired, the CEO recommended focusing its efforts in three key areas (Houpt, 1990; Shumaker, 1990; Winner, Pennell, Bertrand, & Slusarczuk, 1988):

1. Improvements in management/business practices
2. Application of systems engineering techniques
3. Application of computer technology

Management/business practices can be improved through the integration of cooperative, multifunction teams working in a problem prevention mode. These teams

would require a commitment to continuous process improvement, with an emphasis on a balance between product and process development.

Considerations in applying concurrent engineering techniques include the establishment of discipline into the CE process for effective integration of product and process elements, and finding and using synergism among different technical specialists. Additionally, the use of tools and techniques to facilitate CE is of great importance.

The application of computer hardware and software technology is accomplished by modeling, data base management, configuration management, use of computer-based design, computer-based information fusion (trade-off mechanisms), computer-based information presentation (human factors), and computer application architectures (framework).

All three attributes of the CE approach are particularly relevant to the present study:

1. the use of multidisciplinary teams to foster communication between the development and operational communities throughout the design process
2. the application of computer technologies to support the design/development process
3. the application of CE tools (i. e., rapid prototyping)

In fact, Houpt (1990) stressed that the integration of rapid prototyping techniques is an "invaluable tool" in the Integrated Product Development (IPD) or CE environment. This study, in particular, includes the use of rapid prototyping techniques, in just such an environment, to evaluate the HSI.

SECTION 3

Rapid Prototyping

There are two categories of prototypes: engineering prototypes and advance/preproduction prototypes. Engineering prototypes are built to check the feasibility, form, fit, or performance aspects of a particular design, and permit iterative design. Advance/preproduction prototypes are built to near production grade and are used for complete system testing prior to full scale production. Since design changes at this advanced stage are costly, placing more emphasis on the rapid development of engineering prototypes in a CE/ IPD environment (e.g., rapid prototyping) makes sense.

Rapid prototyping, is a CE tool defined by Poindexter (1991) as "any technique that allows for a quick and accurate physical or functional description of a product or process design presented in an easily understandable way." Poindexter (1991) specifically points out that "computer modeling, analysis and simulation tools enable the designers to experiment with several design strategies to assess their unique characteristics and arrive at a 'best' solution to a given product requirement."

Engineering prototypes are required to check feasibility, performance, and tolerances of a design, in addition to, reliability, producibility, maintainability, etc. This process must also include evaluation of the HSI. Houpt (1990) noted that more techniques were available for product prototyping than for process prototyping, and stressed the need for the development of more computer analysis tools oriented toward process design. Furthermore, there was a great need to develop improved rapid prototyping methodology that could utilize a CE/IPD environment and allow the designer to implement quick changes while still

promoting engineering creativity. The development of HSI rapid prototyping falls into this arena.

The integration of rapid prototyping techniques into a CE environment deals with all of the three key areas (facilitating interdisciplinary communication, developing computer design tools, and apply CE procedures). As described by Houpt (1990), the CE or IPD environment is:

> "...one in which a simultaneous approach is taken to engineering a product and its associated manufacturing and support processes. ... A team of personnel made up of different disciplines is responsible for the success or failure of a product... The team fosters open communication so that information normally offered at a later date can be brought out ... earlier, helping to eliminate costly design changes. Increased emphasis is placed on the design of the processes necessary to make the product, and not just the design of the product itself."

Houpt (1990) also mentions that the team uses computer applications for analysis of the product and process design, to provide communication between personnel, and to develop prototypes. Throughout, the focus is on improved efficiency, flexibility, and quality, while at the same time, reducing the product's life cycle cost and time (Houpt, 1990).

Likewise, Shumaker (1990) noted that rapid prototyping in a CE/IPD environment allows one "to develop and demonstrate an improved manufacturing engineering prototyping methodology that concurrently considers product and process design." This would allow engineers to make quick iterations to solve problems, and would not interfere with creativity. Also, there would be an emphasis on fostering effective and efficient communication, since products are designed by teams. This project also required an approach that would "pull together all the key technologies and technical disciplines required to demonstrate the concept of rapid prototyping" (Shumaker, 1990).

Of particular importance, the concept of CE/IPD emphasizes the rapid prototyping of both product and manufacturing processes as early as possible in the weapons system life cycle in order to identify and correct problems early in the acquisition process, thereby making the system more cost-effective. By operating in a prevention mode rather than a correction mode, conceptual design alternatives are demonstrated and evaluated when implementation has the least cost. Shumaker (1990) noted that recent advances in computer-based technologies have increased the prototyping capabilities available, but there still exists a need to develop these capabilities within the CE/IPD environment.

Poindexter (1991) asserts that there is little doubt that "rapid prototyping is one way to help decrease a product's development time and lower its life cycle cost by allowing product and process information to be shared early in the design phase." Further reductions in time and cost can occur if rapid prototyping techniques are applied to CE principles (Houpt, 1990). Computer prototypes can be reviewed and changed quickly thus taking advantage of making improvements early when the cost of change is less. Physical prototypes tend to be costly, but since there is a need to develop prototypes, in an iterative design process that requires multiple prototypes, the only solution is the development of new prototyping techniques. Such is the case with this study, in which the application of simulation to rapid prototyping through use of the SABER (see Section 7) simulation facilities, allows physical iteration for rapid prototyping while also permitting the evaluation of the HSI in a CE environment. This type of simulation is in essence a form of prototyping analogous to the use of factory floor simulation tools. In fact, Shumaker (1990) stated the need for the transitioning of rapid prototyping methodology to other technical areas such as crew systems design and refinement.

SECTION 4

Interface Design

Although referring specifically to software design, Williges, Williges, and Elkerton (1987) noted that design is not a static procedure in that each interface varies according to the application, and that it is iterative, requiring multiple evaluations. The design team should include human factors specialists, as well as computer or other specialists. In other words, design is an iterative process, utilizing rapid prototyping, in a CE/ IPD environment. The design of the HSI for software is readily applied to other processes which involve human-system interfaces.

A flow diagram of the iterative design process for developing interfaces shows three stages:

Stage 1: initial design configuration stage where the interface is specified and the criteria to evaluate the interface are developed, especially with human factors design principles as part of the design objectives

Stage 2: formative evaluation stage of interface deals with techniques for obtaining user feedback to aid the designer in making decisions for design revisions during the iterative design process

Stage 3: summative evaluation of the operational interface design developed by iteration, used to test the final design configuration to ensure that it is functioning properly, which also forms the basis for additional human factors design guidelines.

Williges et al. (1987) noted that the two evaluation stages are similar to "parallel design procedures in other human factors application areas."

In stage one, the initial design stage, areas of importance include compatibility, consistency, memory, structure, feedback, workload, and individualization. The principles as spelled out by Williges et al. (1987) are:

1. Compatibility Principle: Minimize the amount of information recoding that will be necessary
2. Consistency Principle: Minimize the difference in dialogue both within and across various human-computer interfaces
3. Memory Principle: Minimize the amount of information that the user must maintain in short-term memory
4. Structure Principle: Assist the users in developing a conceptual representation of the structure of the system so that they can navigate through the interface
5. Feedback Principle: Provide the user with feedback and error-correction capabilities
6. Workload Principle: Keep user mental workload within acceptable limits
7. Individualization Principle: Accommodate individual differences among users through automatic adaptation or user tailoring of the interface.

With regard to workload, the authors state that user error increases in either overload or underload situations. Therefore, the designer should use any of the various behavioral workload assessment measures when considering various iterations of a design. Furthermore, since “the philosophy of iterative design is to focus the design process on the end user. ... the initial design phase should incorporate user inputs to the greatest extent possible” (Williges, Williges, & Elkerton, 1987). Proper iterative design technique must include evaluation of the HSI if it is to be reliable.

A "structured walk-through" in the initial design phase is more of an informal exercise or evaluation of the information collected and combined during this phase. The modification of design objectives and specifications represents the iteration phase for this stage.

Formative evaluation is the "crux of the iterative design process" (Williges, Williges, & Elkerton, 1987). This is where the interface is implemented, tested, and redesigned until the desired design is achieved, thus allowing the software interface to evolve efficiently. This is best achieved by 1)rapid prototyping procedures and 2)user acceptance testing. (Both of which are implemented in this study.)

Williges et al. (1987) stressed that "...tools are needed to provide rapid development of a candidate interface configuration that can be easily modified after user evaluation." One consideration important to the development of rapid prototyping tools is that the tool must result in a design that is easily modified to facilitate iterative redesign (often providing a simulation of the interface that can be easily converted once the iterative design process is finished).

One approach in the design of software interface is the use of a "facade" for design purposes, and is of particular use in simulating interfaces where the technology is not currently available. Williges et al. (1987) notes that "...user-defined interfaces developed through facades appear to be an important technique to consider during formative evaluation." Simulation has also been used as a means of creating a user-defined interface to create the "illusion of a true interactive session."

The final procedure in formative evaluation, according to Williges, et al., involves some sort of testing that is the "culmination of the iterative design process." This would include questionnaires, as well as performance assessment. In this study, rating scale questionnaires were administered and workload assessments were performed.

In summary, Williges et al. (1987) stated that iterative redesign is the "crux of the iterative design approach in which the results of user acceptance testing define the redesign issues to be addressed in the next design iteration. This redesign is implemented by rapid prototyping and user-defined interface procedures, and continues until the desired design objectives are reached. At this point the formative evaluation process is completed."

These principles, further developed in numerous human factors engineering design guides, support the development of a "usable" HSI conceptual design. Williges et al. (1988) describe "Structured Walk-Throughs" as the final step in an initial design prototyping process. They assert that, once a design concept has been established, "users and designers must then exercise this representation to evaluate its adequacy by taking actual aircraft scenarios and using them to 'walk-through' the control structure of the HCI software to insure successful operation." These authors also suggest that "this step can be accomplished ... by having pilots, controllers, and designers evaluate the conceptual interface through computer-augmented techniques."

Williges et al. (1988) referred to a paper by Gould and Lewis, and noted that there are four critical components for developing and evaluating HCI: early focus on users, interactive design with designers, empirical measurement of learnability and usability of the interface, and iterative design. The concept of focusing on user requirements implies that the design should "incorporate aviation user inputs to the greatest extent possible" (Williges,

Williges, & Fainter, 1988). Williges et al. (1988) state that "as aviation systems become more computer dependent, the emphasis on user-oriented software design must increase in order to improve user acceptance, system efficiency, and safety. Human factors considerations in the design of computer-based aviation systems must address both the issues and the process of software interface design. ...iterative redesign is key to successful HCI software for computer-based aviation systems."

Williges et al. (1988) stressed that there are both benefits and shortcomings to the use of rapid prototyping. Some of the benefits include: the presentation of real interfaces for users to evaluate, a common baseline for users and designers to identify problems, and a practical way to elicit user participation. Some of the shortcomings include: the maintenance of user enthusiasm through several design iterations, the limits of prototyping procedures to represent all features of the interface, the management and control of the iteration process, and the ability to prototype large information systems (Williges, Williges, & Fainter, 1988).

Although much of the rapid prototyping literature deals with the development of the HCI (since computer software now constitutes a major part of the cost and complexity of virtually every modern weapon system), a number of studies have been conducted which included the use of rapid prototyping for crew station design and assessment.

Jones et al. (1981), described a process for integrating the control of flight and navigation subsystems in expectation of achieving "reduced size and complexity or improved performance, or both." Suiter and Sharp (1988) specifically describe the need to apply rapid prototyping to the crew station design process. They noted that prototypes can "aid in acceptance and selling of concepts" and, further, that "they [prototypes] can reveal design

flaws and shortcomings, test interfaces, support iterative man-machine interface design and help manage complexity."

Tamanaha and Bourgeois (1990) presented a methodology for applying rapid prototyping to command and control systems. Moore and Moore (1985) described a system for the rapid prototyping of a pilot-vehicle interface. In their recent study, Kuperman and Sobel (1992) described the application of rapid prototyping to the integration of an advanced target acquisition sensor and ATC capability into a multiplace aircraft.

Section 5

SENSOR FUSION AND INTEGRATION

Brief mention should be made regarding the definitions of sensor fusion and the various avionics concepts employed in this paper. As seen in Figure 1, the "building block approach" shows how a single sensor is incorporated into a multiple sensor array.

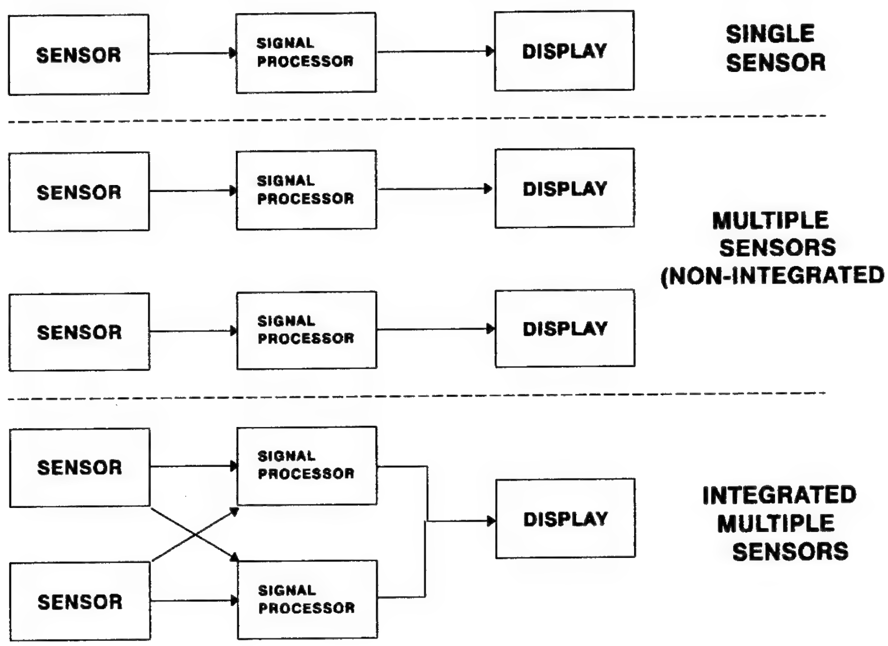

Figure 1. Building Block Approach to Sensor Fusion

This array places the sensors either in parallel (non-integrated, totally separate), or integrated through use of a sensor manager that automatically coordinates the sensors, e.g., slewing the sensors to display a particular target search area after cueing by the operator or an automatic target cuer. The display or image itself from each sensor is processed individually by the image processor. In the next "block" of the approach the sensor images may be "fused" by the signal processor, such that the signals from

each sensor are combined to produce a single display image. While sensor fusion is still largely in the investigative stage, sensor integration is available as a non-materiel solution to multi-sensor avionics concepts.

Luo and Kay (1989) defined multi-sensor integration as "the synergistic use of the information provided by multiple sensory devices to assist in the accomplishment of a task by a system." Multi-sensor fusion is "any stage in the integration process where there is an actual combination (or fusion) of different sources of sensory information into one representational format" (Luo & Kay, 1989) and includes fusion of information from a single sensor that is acquired over time. This separates the integration of multiple sensors at the system and control level from the "actual" fusion of sensor information. They give the example of integrated systems where information from one sensor guides the operation of other sensors, but does not fuse the information. In other words, interaction is not fusion.

Luo and Kay (1989) suggest that potential advantages of integrating and/or fusing information include:

1. information obtained more accurately
2. features otherwise not seen by single sensors
3. less time
4. less cost

Furthermore, they add that these advantages correspond to the following beneficial design attributes:

1. Redundancy: redundant information reduces uncertainty and increases the accuracy of system perception and increases reliability in case of sensor error or failure.

2. Complementarity: complementary perception of features not seen by other sensors
3. Timeliness: processing parallelism possible, difference in speed of single sensors
4. Cost: information costs less when compared to equivalent information with single sensor.

SECTION 6

Simulation

Chiles and Allusi (1979) noted that simulation is the best means of evaluating workload in that it had the greatest "potential of generating large amounts of very useful information on operator workload." However, they mentioned four potential drawbacks that must be considered when using simulation. First, that trained personnel are necessary since complex devices cannot be learned well enough in the short periods of time generally available. Next, since simulators are designed only to simulate the system, there may be some impact on measures due to the unreliability of simulator accuracy, especially in older systems. Furthermore, in simulating a system where many functions are interconnected, there may be difficulty in separating functions for measurement (e.g., performance measures on a simulator evaluate performance of the man—task or concept of primary interest performance and concurrent tasks or concepts performances, the machine, the man-machine interface, and all the possible interactions of these factors). Finally, by emulating one particular system, generalization to other systems that may have different characteristics (e.g., avionics layout or operating procedures) may not be possible.

In this study, these drawbacks have been minimized by using trained personnel that are familiar with the basic avionics concepts, such that familiarization with the simulation is readily achieved. The simulator used has advanced capabilities, and the study has been designed to measure task workload in integrated sensor technology, and not performance or interactions associated with performance. Although generalization may be difficult at this time, multi-sensor technology, as demonstrated in this simulation, could be applied to current

aircraft as a reasonable and cost-saving “off-the-shelf” technology application, which will differ little, if at all, from the simulated concept.

SECTION 7

Methods

Facility

This study was performed using the Strategic Avionics Battle Management Evaluation Research (SABER) laboratory, a facility of the Human Engineering Division, Armstrong Laboratory, Wright-Patterson Air Force Base, Dayton, Ohio. The SABER facility supports human centered research through man-in-the-loop simulation. This facility, while embodying advanced bomber capabilities, is a highly flexible and modular simulator, allowing relatively rapid reconfiguration which can be used to evaluate avionics that may be incorporated into other forms of aircraft, as well. Only one of the four crew stations was configured for this particular study (Figure 2).

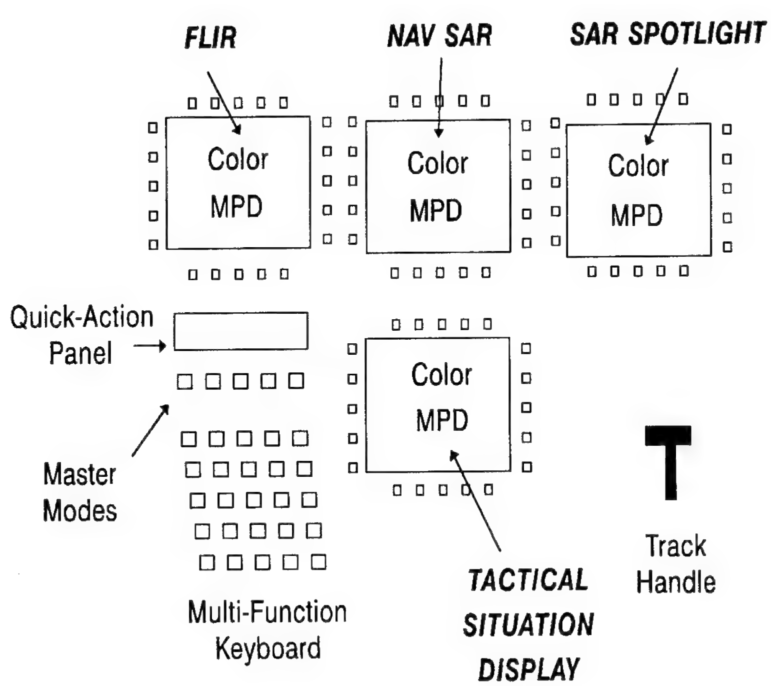

Figure 2. SABER Crew Station

Synthetic aperture radar (SAR) imagery was created based on the planned mission scenarios, and actual Forward Looking Infrared Radar (FLIR) imagery was incorporated after it was digitized from video by the Visual Image Processing Enhancement and Reconstruction (VIPER) facility. Each image(s) of one particular sensor type was displayed on its own multi-purpose display. Images from differing sensors, if applicable, were displayed simultaneously, but on separate displays.

Simulated SAR images were created via a radar model developed at the University of Kansas (Geaga, 1985; Komp, Frost, & Holtzman, 1983). The radar image simulation model was a hybrid of approaches used in previous studies (Kuperman, Wilson, & Perez, 1988). Radar parameters were in accordance with known fundamental principles (Geaga, 1985; Komp, Frost, & Holtzman, 1983), while the geometric factors of terrain masking and vegetative obscuration were based mostly on digital terrain elevation and feature analysis data bases. The area modeled was forested mountains, typical of North Korea. Features (grass, bushes, trees, rivers) were quite similar to actual X-band SAR imagery.

FLIR imagery was digitized in the VIPER facility, by capturing images from actual FLIR video using the Microtime Video Signal Synchronizer 2525. The captured image was interfaced with an International Imaging Systems Model 75 image array processor (I^2S), which allowed for the processing, enhancement, digitization, and hard copying (Polaroid) of the captured images. Imagery presentation and collection for storage was through a Digital Equipment Corporation PDP 11/34 host computer. The original FLIR video contained both live and freeze frame imagery, recorded during flights demonstrating sensor fusion/integration (Kuperman, 1992). A FLIR Systems Incorporated Model 2000B FLIR sensor was used for those flights.

Subjects

Fourteen Subject Matter Experts (SMEs) from the Wright-Patterson Air Force Base population of rated Air Force officers volunteered for this demonstration. The SMEs were 11 males and 3 females, with an average age of 30 years (range 23 - 45). No SME had prior experience with actual SAR sensor imagery. Five SMEs had experience with FLIR imagery, and one additional SME had other non-FLIR infrared experience. The flight experience varied, with a mean of 885 flight hours (range 200-4100 hours).

Overview

The study was divided into three sections. Upon arrival, SMEs were given a brief description of the experimental procedures and were asked to read the consent form, ask questions, and sign the consent form if they wished to participate. A brief background questionnaire was also completed at this time. Ground school was then conducted. This included a detailed briefing as well as training and simulator practice. This was followed by the actual tasking/mission runs with data collection. A debriefing was given afterwards.

Mission Scenarios

Sensor mediated target acquisition must often be performed under operational conditions that are best described as "time compressed." The crew must gain an orientation to the target area (i. e. achieve situational awareness), perform an area search, detect possible targets, assess the detected target(s), correctly recognize the target of interest while rejecting "false alarms," and then, execute the attack, all within a period of perhaps only several seconds. The operational rules of engagement (ROEs, i. e., the conditions under which a target could be attacked) emphasize the need to attack only the correct target in order to avoid fratricide, collateral damage, and/or wasted weapons. These tasks include high sensory and cognitive loading components. Mental effort and time stress are significant components of crew workload during target acquisition. Further, since high value targets are defended, a significant risk element is also involved, resulting in elevated psychological stress. The mission scenarios selected for this study were designed to duplicate these tasks in order to assess these components of crew workload during target acquisition with differing avionics.

A part-mission scenario was developed to provide an operational context to assess the workload involved in each task. The initial task was to perform a standard navigation update at a fix point using Synthetic Aperture Radar (SAR) imagery. The operators then flew the preplanned flight path and viewed the search area with the avionics sensor(s) designated for that particular run for target detection and identification. Based on the sensor information, they designated (accept/reject) as to whether or not a true target was present. Each search area contained three true targets, and one false or empty look (no true target present). Upon designation of a true target, weapons assignment (attack) was automatic.

Avionics Alternatives

The avionics alternatives selected were based on current aircraft capabilities with little or no additional avionics added. Likewise, only sensors that would fit in with the planned operational environment and permit using known ROEs were selected (e.g. laser line scanner was not used since this would require a direct overfly of the target for reconnaissance missions, and would require an "over the shoulder" weapons delivery after target selection).

Using a baseline of the lower resolution synthetic aperture radar (SAR) available for both the navigation update and the initial target acquisition task, progressive integration of additional avionics occurred with each subsequent mission from which to compare perceived changes in workload (repeated measures design) as noted in Table 1.

MISSION RUN	AVIONICS CONCEPT ALTERNATIVE (Independent Variables)
1	NAV SAR + SAR/HRGM
2	RUN 1 + SPOTLIGHT SAR
3	RUN 2 + ATC
4	RUN 3 + FLIR

NAV : Navigation SAR : Synthetic Aperture Radar

HRGM : High Resolution Ground Map

ATC: Automatic Target Cuer FLIR: Forward Looking Infrared

Table 1. Mission Run: Avionics Alternatives

RUN 1 (NAV SAR - SAR/HRGM): Several current aircraft (F-15E, B-1B, B-2) already possess a synthetic aperture radar which is used to support radar fix taking for navigation and weapon delivery purposes. SAR provides a day/night, all weather imaging capability, with a resolution on the order of a few meters. Little or no impact on crew training or workload is expected since the SAR would be used in essentially the same manner that it is currently employed. NAV SAR and SAR/HRGM are essentially the same modality; NAV SAR referring to navigation usage, SAR/HRGM for target acquisition usage.

For each run, the operator entered the coordinates of the suspected target location and radar fix point for the navigation update into his navigation computer, then flew to an appropriate range and bearing with respect to those coordinates to perform the navigation update. The SME then initiated the NAV SAR map and performed radar scope interpretation (RSI) on the navigation update fix point (antenna). After locating the fix point on the SAR image display, he manually designated the fix location by moving a designation crosshair over the fix point on the SAR image. The fix point was accepted to update the computer automatically, or rejected (operator repeats navigation update sequence).

The operator then flew to an appropriate range and bearing with respect to those coordinates entered for the suspected target location. He then initiated a SAR map and performed radar scope interpretation (RSI). After detecting the target on the SAR image display, he manually designated the target location by moving a designation crosshair over the detected target on the SAR image. The suspected target was accepted as a true target or rejected as a false alarm. If accepted as a true target, the target location was handed off to the weapon delivery system (Stores Management System), for preplanned automatic weapon delivery.

The ROEs for this study permitted attacking the target based solely on the capability to perform detection, although this might violate requirements for fratricide avoidance and minimizing collateral damage. Despite this deficiency, this concept provided the study with a current capability and well understood set of crew tasks for purposes of comparison. Furthermore, the navigation update performed at the beginning of each mission run provided a baseline task for a repeated measures comparison.

RUN 2 (NAV SAR - SAR/HRGM + SPOTLIGHT SAR): This alternative added a spotlight or higher resolution imaging mode to the baseline SAR. The navigation SAR mode (resolution in meters) was retained for (manual) target detection. The same crosshair placement and designation schema of the previous run was retained. In this concept, however, each target designation action on the navigational SAR map results in commanding a higher resolution, "spotlight" SAR map of the designation location. Spotlight SAR mode initiation is "automated" in that automated radar antenna pointing and radar mode control are functions of an implied Sensor Management Subsystem (SMS). It is assumed that a 2 to 4 times improvement in image resolution supported a higher level of information extraction during RSI/target identification.

After spotlight SAR designation, the navigation SAR image was retained on a display to provide contextual information and support situational awareness. The operator then performed RSI on each spotlight image and repeated the target designation/rejection process as in the previous run.

This concept supported identification (a more stringent ROE) but also introduced additional crew tasks (which may result in elevated crew workload). Additional training in

RSI of spotlight SAR imagery would be required, but attacking false alarms (i. e., non-targets) would probably be reduced.

RUN 3 (NAV SAR - SAR/HRGM + SPOTLIGHT SAR + ATC): An automatic target cuer (ATC) is integrated with the SAR. The ATC automatically processes the navigation SAR image, performs the target detection and commands the initiation of the spotlight mode images for each detected target. The navigation SAR map, followed by the appearance of the spotlight SAR map, was overlaid graphically to depict detection locations and presented to the operator. The crew member still performed manual RSI on each spotlight image followed by target designation/rejection.

ROEs would not be expected to be affected by this concept. Some reduction in crew workload and time demands is expected by automating the detection task with navigation SAR ATC. However, pilot acceptance and workload shifting are variables which may counter the expected reduction in workload.

RUN 4 (NAV SAR - SAR/HRGM + SPOTLIGHT SAR + ATC + FLIR): Forward looking infrared radar was added to exploit the thermal signature (reflected or emitted) of the target. It may complement SAR by differentiating between decoys and actual targets. It is effective at night as well as during the day, but is susceptible to adverse weather. It is also range-limited to a few NM (SAR can be employed at much greater ranges in 10s of NM). The boresight of FLIR, however, can be slewed in azimuth and elevation to image ground points which are not directly on the flight path (and even behind) such that the field of regard of the FLIR is the full lower hemisphere.

Since the FLIR image would only be commanded against a detected potential target, a static FLIR image, centered on that location, was presented to the crew for target

detection/identification along with the other SAR maps as in the previous run. FLIR pointing and mode control were implied functions of the SMS.

The ATC automatically processes the navigation SAR image, performs the target detection and commands the initiation of the spotlight mode and FLIR images for each detected target. The navigation SAR map, followed by the appearance of the spotlight SAR map and the FLIR imagery was presented to the operator as in the previous run. The crew member still performed manual RSI and either designated identified targets to the weapon delivery subsystem or rejected false alarms.

By providing a capability to perform multiple sensor target identification which exploits a totally distinct portion of the electromagnetic spectrum and providing additional capability to differentiate between decoys and actual targets, FLIR would permit ROEs that demand higher confidence weapon releases. This would result in minimizing fratricide risks, collateral damage, and wasted weapons (false alarm reduction).

Additional image interpretation training would be required for the crew to fully exploit the additional sensor. Workload might be expected to increase due to the "switchology" of the second sensor. And, since an ATC and SMS are already in place, workload and time demands should be reduced, while exploiting the infrared imagery. However, pilot acceptance and workload shifting are variables which might offset the expected reduction in workload. Furthermore, the requirement to interpret imagery from an additional sensor may also increase the perceived workload.

SECTION 8

Analysis and Results

Workload

The Subjective Workload Assessment Technique (SWAT) card sorts were subjected to standard routine analysis, (Reid, Potter, & Bressler, 1989) for scale development. The Kendall's Coefficient of Concordance for the group was 0.76. A value of 0.75 and above indicates a relatively homogenous group, allowing a group scaling solution to be used for finding the rescaled value during data analysis. Since the value was near-borderline, a prototyped scaling solution which rescales the values depending upon the individual's prototype, a more conservative scaling approach, was chosen. Prototyping groups the subjects based on their perceptions of the relative importance of the three dimensions in SWAT (i. e., a "time" prototype indicates that the subject considers the Time Load dimension to be the dominant contributor to the perceived workload). The prototype analysis showed that three subjects were time prototypes, three were effort prototypes, and the remaining eight subjects were stress prototypes. The Kendall's Coefficient of Concordance for each prototype was 0.96 for the time prototype, 0.94 for the effort prototype, and 0.89 for the stress prototype. Evaluation of the axiom violations verified acceptability for prototype scaling.

SWAT event scores (time, effort, stress triplets) were elicited at the completion of the navigation update and of each segment of the mission for the specific avionics concept demonstrated (repeated measures design). Using the prototype scales, the event scores were converted into SWAT values in the range of zero to 100. (Values in the range of 30 to 50

are considered to be indicative of a potential workload saturation condition as noted by Reid and Colle (1988).

SWAT values for each avionics concept were then subjected to standard statistical analyses. A one-way analysis of variance was conducted to investigate the possible difference in workload between the five tasks (navigation update and target acquisition using each of the four avionics concepts). Multiple paired comparison t-tests were used to identify the specific tasks which gave rise to statistically significant workload estimations. (All tests were performed at a confidence level of $p < 0.05$.) The means and standard deviations ($n = 14$) of the SWAT values for the five tasks are shown in Table 2, below.

Task	Mean SWAT Values	Standard Deviation
Navigation Update	2.54	4.58
SAR/HRGM	13.86	16.41
+ Spotlight SAR	14.11	13.00
+ ATC	14.48	12.57
+ FLIR	20.91	15.03

Table 2. SWAT values

The ANOVA table generated from the data above is shown in the table below.

SOURCE	DF	SS	MS	F-ratio	P
Task	4	2463.22	615.805	3.65	0.0096
Error	65	10958.27	168.5887		
TOTAL	69				

DF: degrees of freedom SS: sum-of-squares MS: mean square P: probability

Table 3. SWAT ANOVA table

Since the five tasks were found to be statistically different from each other with respect to expected workload (as measured by the mean SWAT values), a family of paired comparison t-tests was performed to identify the specific difference(s). The mean of the navigation update task was found to be statistically lower than the means of any of the other four tasks. None of the other inter-task comparisons was found to be statistically significant. Figure 3 presents these findings graphically. Additional comparisons made of subgroups based first on prototype, then on flight experience, and finally, on sensor experience, were also found to be not statistically significant.

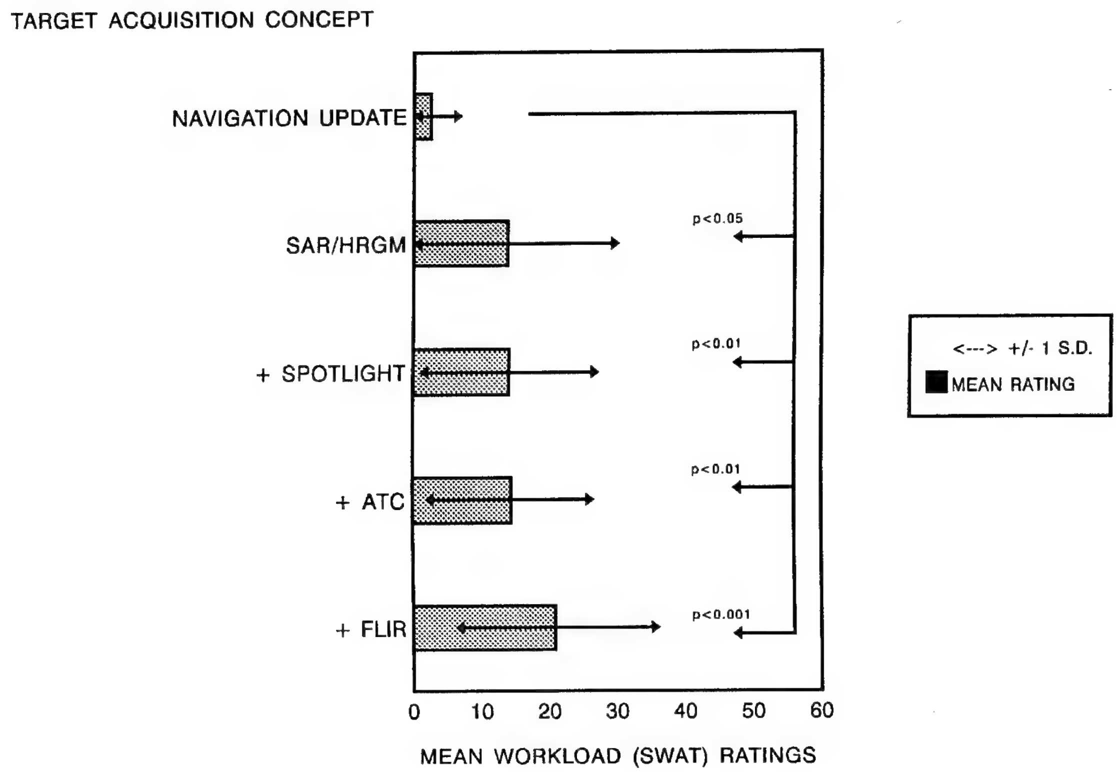

Figure 3. Family of Paired t-Tests Inter-Task Comparisons

The mean SWAT values for the tasks were all less than 30. This is indicative of an acceptable level of workload. Although some individual ratings, most notably with addition of the FLIR, were in the 30-40 range (30-50 is indicative of potential workload saturation), none were in the workload saturation range (>50). Since SWAT is best at evaluating workload in the less than 50 range, the values can be considered valid for this study.

The magnitude of the variance merited further analyses. In an attempt to better understand the large variances, the subjects were divided into subgroups according to their SWAT prototype, sensor experience, and flight experience. No statistically significant between group difference was found for any of these subgroup divisions. However, the size of many of the subgroups (n=3) may have been too small to reliably reflect any differences. Since practice has been shown to reduce response variability, inadequate training may have

contributed to some of the variability. Reduced variance might have been obtained by increased training in the workload reporting task. Additionally, individual inconsistencies may have exacerbated variability due to axiom violations which are corrected for during the prototyping analysis.

As expected, the Navigation Update task had intrinsically lower workload than the other tasks which incorporated enhancements to the avionics capabilities. Although no statistically significant difference was shown between any of the four tasks, a trend of increasing mean values of workload was observed with each incremental enhancement.

Two observations can be made in comparing these SWAT values for the target acquisition tasks. First, run 2 and run 3 of the avionics concepts differ from each other only with respect to the incorporation of an ATC. The ATC automates the detection of possible targets, presumably reducing operator workload. Since these tasks were not found to differ in workload, this might be attributed to full task workload being dominated by the imagery interpretation component of the task; automation of the detection component does not, then, significantly reduce perceived full task workload.

The second observation relates to run 4, which introduced a FLIR sensor. This addition resulted in elevated workload which, while not statistically significant, is at least in agreement with the conjecture of increased workload related to imagery interpretation.

Operator Interface Evaluation

Three rating scale questions were administered at the conclusion of the SABER avionics concept demonstrations. They probed the SME's overall impressions of the experiment. The first question explored the realism of the "mini-scenario" used to provide an operational context, the second dealt with the adequacy of the "ground school" pre-briefing, and the third question related to the realism of the mission tasking.

Figure 4 presents the mean SME ratings for the general areas of consideration.

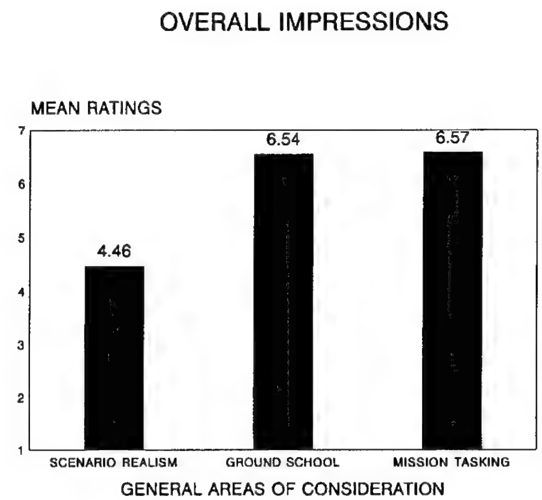

Figure 4. Mean SME Rating for General Areas of Consideration

As can be seen from the Figure, the SMEs were generally neutral with regard to scenario realism but were very positive in their ratings of ground school preparation and mission tasking.

Avionics Concept Evaluations

A series of rating scale-type questions were administered to the SMEs during post-concept demonstration debriefing. Four avionics concepts for air-to-ground target acquisition were actually demonstrated to the SMEs in the SABER simulation runs. They were:

1. Manual target detection, location, and identification using the navigation SAR mode of the multimode radar (NAV SAR)
2. Manual target detection with NAV SAR and target identification using a (new) higher resolution Spotlight SAR radar mode (+ SPOTLIGHT)
3. Automated target detection through the addition of an automated target cuer exploiting the NAV SAR imagery and manual identification using the Spotlight mode images (+ ATC)
4. The same as concept #3, but with the addition of a static FLIR image to further support target identification (+ FLIR)

A fifth concept was described (but not demonstrated) to the SMEs during their post-simulator debriefing. This concept eliminated the NAV SAR ATC (- ATC) but retained the SPOTLIGHT SAR mode and the FLIR image, as in concept #4. (Each rating scale probe was followed by an opportunity for the SME to document supporting impressions or information. These responses are documented in the Appendix.)

Each of these concepts, as implemented in the SABER simulator, was explored by means of a rating scale questionnaire. The capability of the concept, including the SABER crew-vehicle interface, was explored with regard to:

1. Support of Target Detection
2. Support of Target Location
3. Support of Target Identification
4. General Impression of Associated Workload
5. Expected Impact on the Training Support System
6. Crew Confidence in Employing the Avionics
7. Support to Situational Awareness
8. Support of Mission Pacing
9. Likelihood of Generating False Alarms
10. Contribution of the Sensor Management Subsystem
11. Support of Operator-in-the-Loop Control

An identical rating scale questionnaire was administered immediately following the demonstration of the specific avionics concept being evaluated. Each of the ratings from each SME and for each avionics concept was analyzed by means of a one-way ANOVA to determine if any of the concepts were statistically superior in providing/supporting the attribute. Where statistical significance was found, a family t-Test ($p < 0.05$) was conducted to identify the specific concept(s) expected to provide the superior quality.

Detection:

Figure 5 presents the mean SME ratings elicited with respect to the capability of the SABER HSI to support the aircrew in performing target detection.

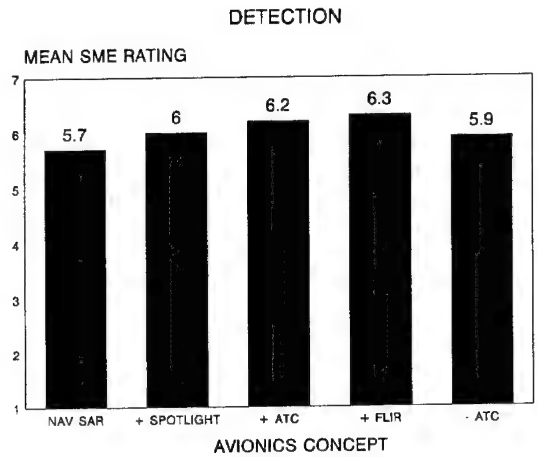

Figure 5. Mean SME Ratings for Detection

The ANOVA revealed no statistically significant differences between the avionics concepts with regard to their relative capabilities to support the function of target detection. Since the mean ratings were all greater than 5.7, each of the avionics concepts can be considered to be fully capable of supporting the aircrew in performing target detection.

Location:

Figure 6 presents the mean SME ratings elicited with respect to the capability of the SABER HSI to support the aircrew in performing target location. Again, the ANOVA revealed no significant differences between the five concepts. The minimum mean rating for this HSI attribute was 5.6 and all concepts can be described as fully capable of supporting the operator in performing target location.

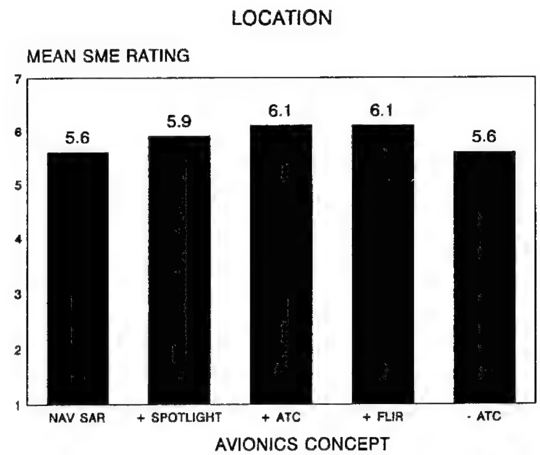

Figure 6. Mean SME Ratings for Location

Identification:

Figure 7 presents the mean SME ratings for target identification.

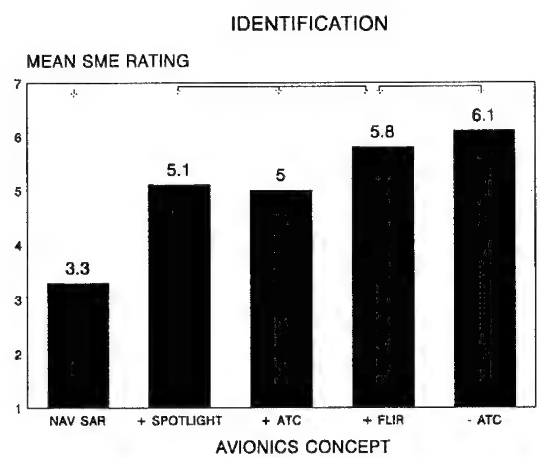

Figure 7. Mean SME Ratings for Identification

In the case of identification, the ANOVA revealed significant differences ($p < 0.01$) between the five avionics concepts under study. Table 4 presents results of the family t-Test.

CONCEPT			
NAV SAR	A		
+ SPOTLIGHT		B	
+ ATC		B	
+ FLIR		B	C
- ATC			C

TABLE 4. Results of Family t-Test for Identification

As can be seen from this Table, the five avionics concepts divided into three distinct groups (A, B, and C) with respect to SME judgment as to their respective capabilities to support target identification. The NAV SAR concept was significantly different from each of the other concepts. The + SPOTLIGHT, + ATC, and + FLIR concepts, as a group, were different from either the NAV SAR or - ATC concepts. The + FLIR and - ATC concepts, as a group, were different than the three other concepts.

Workload:

Figure 8 presents the mean SME ratings for their expectation with regard to aircrew workload. The ANOVA indicated that there was a significant difference ($p < 0.01$) between the concepts. Table 5 presents the results of the family t-Test for the ratings on the workload attribute.

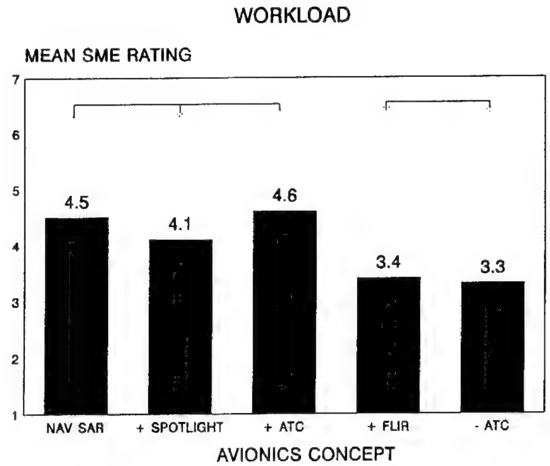

Figure 8. Mean SME Ratings for Workload

CONCEPT		
NAV SAR	A	
+ SPOTLIGHT	A	
+ ATC	A	
+ FLIR		B
- ATC		B

TABLE 5. Results of Family t-Test for Workload

As is indicated in the Table, the NAV SAR, + SPOTLIGHT, and + ATC concepts, as a group, were rated as having lower workload associated with them than were the + FLIR and - ATC concepts, again as a group. (This finding is in general agreement with the results of the SWAT workload measurement in which NAV SAR was found to support a lower level of workload than the other four target acquisition concepts.)

Training:

Figure 9 presents the mean SME ratings elicited with regard to the impact of

each avionics concept on the training support system. The ANOVA revealed no difference between the concepts with regard to this attribute.

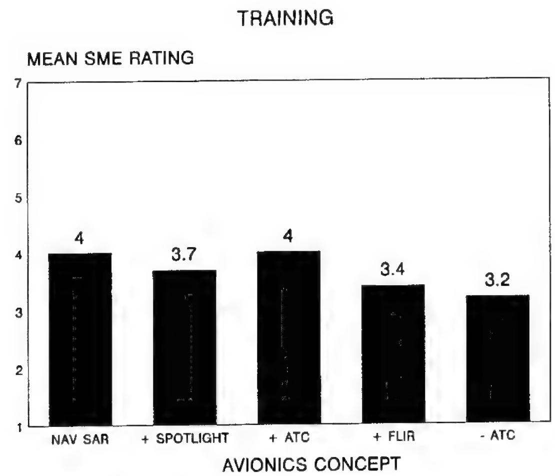

Figure 9. Mean SME Rating for Training

Confidence:

Figure 10 presents the mean SME ratings for crew confidence. No significant difference between the five concepts was identified by the ANOVA.

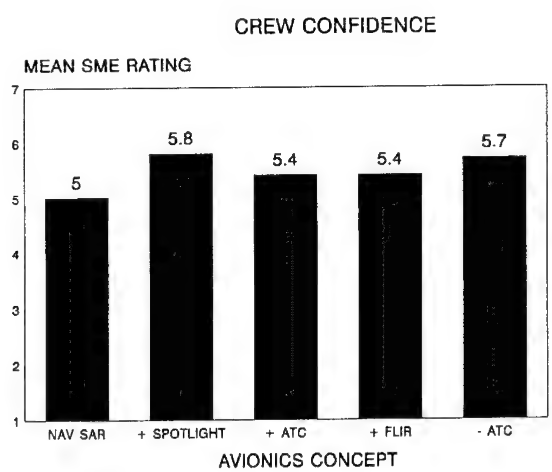

Figure 10. Mean SME Ratings for Crew Confidence

Situational Awareness:

Figure 11 presents the mean SME ratings for crew situational awareness. No significant difference between the five concepts was identified by the ANOVA.

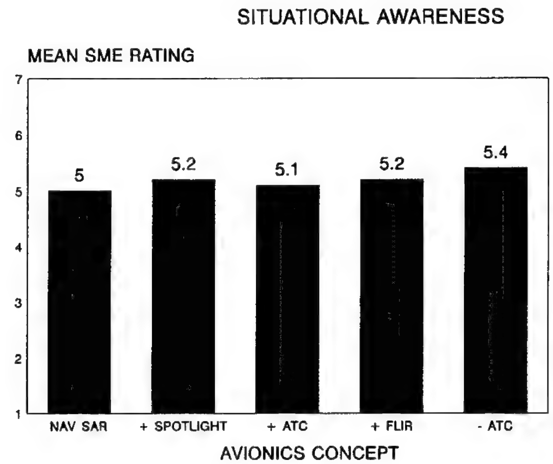

Figure 11. Mean SME Ratings for Situational Awareness

Mission Pacing:

Figure 12 presents the mean SME ratings for the expected capability of each of the avionics concepts to support the aircrew in performing mission pacing. The ANOVA did not identify the existence of any significant differences between the concepts.

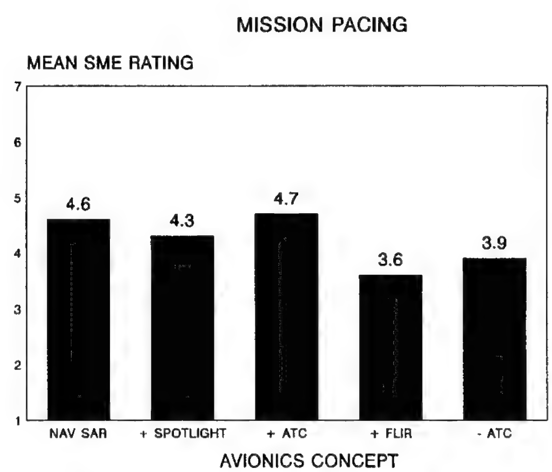

Figure 12. Mean SME Ratings for Mission Pacing

False Alarms:

This rating scale question explored the SMEs expectation that a target acquisition avionics capability would result in the occurrence of false alarms (i. e., non-target incorrectly declared to be targets). The ANOVA revealed a significant difference ($p < 0.05$) between the five concepts. Figure 13 presents the means of the SME ratings and Table 6 depicts the results of the *post hoc* family t-Test.

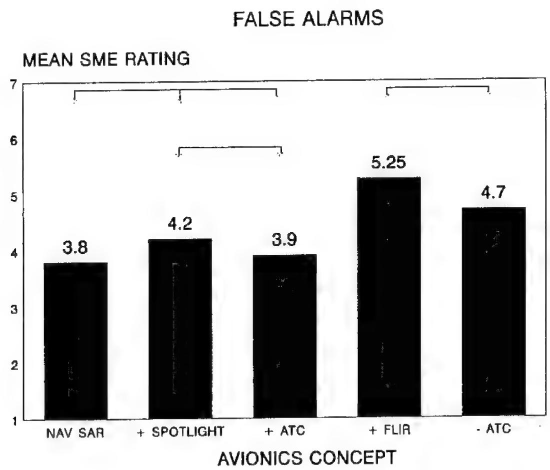

Figure 13. Mean SME Ratings for False Alarms

CONCEPT			
NAV SAR	A		
+ SPOTLIGHT	A	B	
+ ATC	A	B	
+ FLIR			C
- ATC			C

TABLE 6. Results of Family t-Test for False Alarms

This result is interesting in that it supports the contention that addition of a second sensor modality (i. e., the FLIR) would reduce the occurrence of false alarms.

Sensor Manager:

Two of the avionics concepts (+ ATC and + FLIR) include an automated capability to point a second sensor or sensor mode at the location of each suspect target location (i. e., each detection). This capability would be provided by a sensor manager subsystem. The ANOVA revealed no difference between these two avionics concepts in terms of the expected contribution to mission effectiveness by the sensor manager. Figure 14 presents the mean SME ratings for this capability.

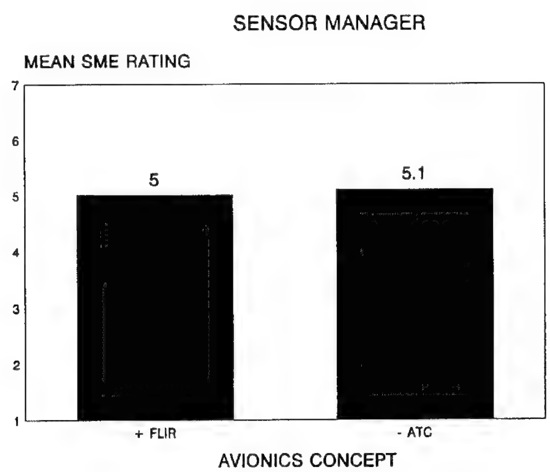

Figure 14. Mean SME Ratings for the Sensor Management Subsystem

Man-in-the-Loop:

This attribute referred to the ability of an avionics concept to keep the operator involved in the process (in this case, target acquisition) taking place. Figure 15 presents the mean SME ratings for man-in-the-loop involvement.

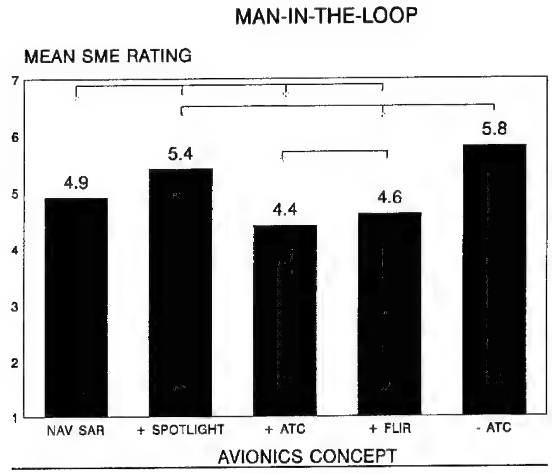

Figure 15. Mean SME Ratings for Man-in-the-loop Involvement

The ANOVA revealed a significant difference between the concepts ($p < 0.01$). Table 7 depicts the results of the family t-Test employed to examine this difference.

CONCEPT			
NAV SAR	A		
+ SPOTLIGHT	A	B	
+ ATC	A		C
+ FLIR	A	B	C
- ATC		B	

TABLE 7. Results of Family t-Test for "Man-in-the-Loop"

As shown in the Table, the concepts fall into three overlapping groups. The most interesting observation to be drawn from these results is that the two concepts which included the ATC (+ ATC and + FLIR) resulted in lowest rating with respect to a feeling of "in-the-loop control."

SECTION 9

Conclusions

The first part of this study showed the effect of progressive integration of multiple sensors and automatic target cueing for target acquisition on subjective workload. The use of SWAT was readily acceptable to the subjects and easily utilized for this type of study. Workload, as estimated by the mean SWAT values, was found to be within acceptable ranges (i. e., mean SWAT was < 30) for all tasks.

All target acquisition tasks, representing incremental integration of enhancements in avionics capabilities, were found to be of higher intrinsic workload than the navigation update task. No statistically significant difference in workload was found between any of the four target acquisition tasks, although a trend of increasing mean values was seen with each successive integration of enhanced avionics.

Further integration of additional sensors will, in all probability, increase workload to some extent. This potential increase in workload might not be offset by decrements in perceived workload associated with automation and enhanced target identification capabilities, as indicated in this study. Thus, it is reasonable to postulate that further integration of additional sensors might produce a potential overload situation. Future avionics which integrate multiple sensors will need to allow for this possibility during the initial stages of the design process. Additionally, it is recommended that future evaluations continue to use SWAT as a measure of subjective workload.

The study also demonstrated the utility of this form of simulation within the context of a CE environment. The ratings by the SMEs verified that the simulation was adequate, and that the mission tasking and preparation was a positive experience.

Each concept's HSI, evaluated by questionnaire, showed that each of the concepts could support target detection and location. Identification, however, was aided by the integration of additional sensor modalities. The occurrence of false alarms was also expected to be decreased by the integration of additional sensors. The ATC was not expected to aid in the identification process, nor to be of help in decreasing false alarms.

Workload evaluations seemed to indicate a slight increase associated with the addition of the FLIR. Although the avionics concept which included the FLIR was not found to be significantly different from the three other target acquisition concepts, it did result in the highest SWAT score (mean SWAT = 20.91). The rating scale which explored workload received the second lowest rating (i. e., highest workload). In none of the cases, however, was workload judged to be so high as to be of concern.

Each of the concepts did not significantly impact training, confidence, situational awareness, or mission pacing. Of concern, is that not only were the sensor management subsystem and ATC perceived as not contributing to mission effectiveness, but that the addition of the ATC resulted in a loss of feeling of "in-the-loop control." As more systems are integrated, with an anticipated increase in automation, the effects of automation on the HSI will need to be further evaluated and addressed. This can be accomplished best in a CE environment utilizing rapid prototyping of the HSI early in the design and acquisition process, as demonstrated here.

REFERENCES

Adelman, L., & Donnell, M. L. (1988). An Empirical Study Comparing Pilots' Interrater Reliability Ratings for Workload and Effectiveness. *IEEE Transactions on Systems, Man, and Cybernetics*, 18(6), 978-981.

Bortolussi, M. R., Kantowitz, B. H., & Hart, S. G. (1986). Measuring Pilot Workload in a Motion Base Trainer. A Comparison of Four Techniques. *Applied Ergonomics*, 17(4), 278-283.

Byers, J. C., Bittner, A. C., Jr., Hill, S. G., Zaklad, A. L., & Christ, R. E. (1988). Workload Assessment of a Remotely Piloted Vehicle (RPV) System. *Proceedings of the Human Factors Society- 32nd Annual Meeting*, 1145-1149.

Casali, J. G., & Wierwille, W. W. (1983). A Comparison of Rating Scale, Secondary-Task, Physiological, and Primary-Task Workload Estimation Techniques in a Simulated Flight Task Emphasizing Communications Load. *Human Factors Society Bulletin*, 25(6), 623-641.

Chiles, W. D., & Alluisi, E. A. (1979). On the Specification of Operator or Occupational Workload with Performance-Measurement Methods. *Human Factors Society Bulletin*, 21(5), 515-528.

Corwin, W. H. (1992). In-Flight and Postflight Assessment of Pilot Workload in Commercial Transport Aircraft Using the Subjective Workload Assessment Technique. *The International Journal of Aviation Psychology*, 2(2), 77-93.

Eggemeier, F. T. (1981). Current Issues in Subjective Assessment of Workload. *Proceedings of the Human Factors Society- 25th Annual Meeting*, 513-517.

Eggemeier, F. T., Biers, D. W., Wickens, C. D., Andre, A. D., Vreuls, D., Billman, E. R., & Schueren, J. (1990). *Performance Assessment and Workload Evaluation Systems: Analysis of Candidate Measures*. HSD-TR-90-023, Human Systems Division, Brooks Air Force Base, Texas.

Gawron, V. J., Schifflet, S. G., & Miller, J. C. (1989). Measures of In-Flight Workload. In R. S. Jensen (Ed.), *Aviation Psychology* (pp. 240-287). Brookfield: Gower Technical.

Geaga, J. V. (December 1985). *Synthetic Aperture Radar Target and Terrain Simulator*. Report No. 86-4R, Autonomous Systems Laboratory, Northrop Research and Technology Center, Palos Verdes Peninsula, California.

Gopher, D., & Donchin, E. (1986). Workload—An Examination of the Concept. In K. R. Boff, L. Kaufman, & J. P. Thomas (Eds.), *Handbook of Perception and Human Performance* (Vol. 2, pp. 41-1 - 41-49). New York: John Wiley and Sons.

Hart, S. G., Childress, M. E., & Bortolussi, M. (1981). Defining the Subjective Experience of Workload. *Proceedings of the Twenty-fifth Annual Meeting of the Human Factors Society*, 527-531.

Hill, S. G., Iavecchia, H. P., Byers, J. C., Bittner, A. C., Jr., Zaklad, A. L., & Christ, R. E. (1992). Comparison of Four Subjective Workload Rating Scales. *Human Factors Society Bulletin*, 34(4), 429-439.

Houpt, T. J. (1990). *The Importance of Rapid Prototyping to Manufacturing and Integrated Product Development.* White Paper. Concurrent Engineering Office, Manufacturing Technology Directorate, Wright Research and Development Center, Wright-Patterson Air Force Base, Ohio.

Jex, H. R. (1988). Measuring Mental Workload: Problems, Progress, and Promises. In P. A. Hancock & N. Meshkati (Eds.), *Advances in Psychology: 52. Human Mental Workload* (pp. 5-39). New York: North-Holland.

Jex, H. R., & Clement, W. F. (1979). Defining and Measuring Perceptual-Motor Workload in Manual Control Tasks. In N. Moray (Ed.), *Mental Workload: Its Theory and Measurement* (pp. 125-177). New York: Plenum Press.

Johanssen, G., Moray, N., Pew, R., Rasmussen, J., Sanders, A., & Wickens, C. (1979). Final Report of the Experimental Psychology Group. In N. Moray (Ed.), *Mental Workload: Its Theory and Measurement* (pp. 101-114). New York: Plenum.

Jones, A. L., Mahesh, J. K., Kollodge, M. A., Pledger, D. P., Rang, E. R., & Graham, K. D. (1981). *Integrated Control Design Techniques.* AFWAL-TR-81-3074, Air Force Wright Aeronautical Laboratory, Wright-Patterson Air Force Base, Ohio.

Kalawsky, R. S. (1987). Pilot Integration and the Implications on the Design of Advanced Cockpits. *AGARD Conference Proceedings No. 425: The Man-Machine Interface in Tactical Aircraft Design and Combat Automation*, 24-1 - 24-3.

Kantowitz, B. H., & Casper, P. A. (1988). Human Workload in Aviation. In E. L. Wiener & D. C. Nagel (Eds.), *Human Factors in Aviation* (pp. 157-187). New York: Academic Press, Inc.

Komp, E. E., Frost, V. S., & Holtzman, J. C. (January 1983). *User Manual for the Radar Image Simulator.* RSL TR 581-1, Remote Sensing Laboratory, University of Kansas Center for Research, Inc., Lawrence, Kansas.

Kuperman, G. G. (September 1992). *Operator Interface Assessment for the Sensor Fusion Flight Demonstration Program.* AL-TR-1992-0117, Armstrong Laboratory, Wright-Patterson Air Force Base, Ohio.

Kuperman, G. G., & Sobel, A. L. (1992). Design of the Man-Machine Interface for an Automatic Target Cuer System. *Proceedings of the Institute of Electrical and Electronic Engineers (IEEE) 1992 National Aerospace and Electronics Conference*, Dayton, Ohio.

Kuperman, G. G., Wilson, D. L., & Perez, W. A. (June 1988). *Relocatable Target Acquisition Performance With Simulated Synthetic Aperture Radar Imagery.* AAMRL-TR-88-025, Armstrong Aerospace Medical Research Laboratory, Wright-Patterson Air Force Base, Ohio.

Leplat, J. (1978). Factors Determining Workload. *Ergonomics*, 21, 143-149.

Luo, R. C., & Kay, M. G. (1989). Multisensor Integration and Fusion in Intelligent Systems. *IEEE Transaction on Systems, Man and Cybernectics*, 19(5), 901-931.

Lysaght, R. J., Hill, S. G., Dick, H. O., Plamondon, B. D., Linton, P. M., Wierwille, W. W., Zaklad, A. L., Bittner, A. C., Jr., & Wherry, R. J. (1989). *Operator Workload: Comprehensive Review and Evaluation of Operator Workload Methodologies*. ARI Technical Report 851, Army Research Institute, Fort Bliss, Texas.

Matthes, G. W. (1987). Mission Planning and Proper Design: The Long Range Connection. *AGARD Conference Proceedings No. 425: The Man-Machine Interface in Tactical Aircraft Design and Combat Automation*, 3-1 - 3-4.

Moore, R. D., & Moore, C. A. (1985). Pilot/Vehicle Interface Design System (PIVIDS). *Proceedings, IEEE National Aerospace and Electronics Conference*, Dayton, Ohio.

Moray, N. (1982). Subjective Mental Workload. *Human Factors Society Bulletin*, 24(1), 25-40.

Nygren, T. E. (1991). Psychometric Properties of Subjective Workload Measurement Techniques: Implications for Their Use in the Assessment of Perceived Mental Workload. *Human Factors Society Bulletin*, 33(1), 17-33.

O'Donnell, R. D., & Eggemeier, F. T. (1986). Workload Assessment Methodology. In K. R. Boff, L. Kaufman, & J. P. Thomas (Eds.), *Handbook of Perception and Human Performance* (Vol. 2, pp. 42-1 - 42-49). New York: John Wiley and Sons.

Papa, R. M., & Stoliker, J. r. (1987). Pilot Workload Assessment: A Flight Test Approach. *AGARD Conference Proceedings No. 425: The Man-Machine Interface in Tactical Aircraft Design and Combat Automation*, 8-1 - 8-12.

Poindexter, J. W. (1991). Rapid Prototyping in an Integrated Product Development Environment. *SAE, Aerospace Atlantic Conference*, Dayton, Ohio.

Reader, D. C. (1987). Human Limitations in Flight and Some Possible Remedies. *AGARD Conference Proceedings No. 425: The Man-Machine Interface in Tactical Aircraft Design and Combat Automation*, 5-1 - 5-7.

Rehmann, J. T., Stein, E. S., & Rosenberg, B. L. (1983). Subjective Pilot Workload Assessment. *Human Factors Society Bulletin*, 25(3), 297-307.

Reid, G. B., & Colle, H. A. (1988). Critical SWAT Values for Predicting Operator Overload. *Proceedings of the Human Factors Society- 32nd Annual Meeting*, 1414-1418.

Reid, G. B., & Nygren, T. E. (1988). The Subjective Workload Assessment Technique: A Scaling Procedure for Measuring Mental Workload. In P. A. Hancock & N. Meshkati (Eds.), *Advances in Psychology: 52. Human Mental Workload* (pp. 185-218). New York: North-Holland.

Reid, G. B., Potter, S. S., & Bressler, J. R. (1989). *Subjective Workload Assessment Technique (SWAT): A User's Guide*. AAMRL-TR-89-023, Armstrong Aerospace Medical Research Laboratory, Wright-Patterson Air Force Base, Ohio.

Reid, G. B., Shingledecker, C. A., & Eggemeier, F. T. (1981). Application of Conjoint Measurement to Workload Scale Development. *Proceedings of the Human Factors Society- 25th Annual Meeting*, 522-526.

Sexton, G. A. (1988). Cockpit-Crew Systems Design and Integration. In E. L. Wiener & D. C. Nagel (Eds.), *Human Factors in Aviation* (pp. 495-526). New York: Academic Press, Inc.

Sheridan, T. B. (1980). Mental Workload—What Is It? Why Bother With It? *Human Factors Society Bulletin*, 23(2), 1-2.

Sheridan, T. B., & Simpson, R. W. (1979). Toward the Definition and Measurement of the Mental Workload of Transport Pilots. Final report DOT-OS-70055, Cambridge Massachusetts Institute of Technology.

Shumaker, G. C. (1990). *Integrated Product Development Program Strategy*. White Paper. Concurrent Engineering Office, Directorate of Manufacturing Technology, Wright Research and Development Center, Wright-Patterson Air Force Base, Ohio.

Suiter, J. M., & Sharpe, T. G. (1988). Software, Hardware, and Rapid Prototyping Considerations in Advaned Crew Stations Design. *Proceedings, 8th Digital Avionics Systems Conference*, San Jose, California.

Tamanaha, D. Y., & Bourgeois, P. J. (1990). Rapid Prototyping of Large Command, Control, Communications, and Intelligence C3I Systems. *Proceedings, IEEE Aerospace Applications Conference*, Vail, Colorado.

Tsang, P. S., & Johnson, W. W. (1989). Cognitive Demands in Automation. *Aviation, Space, and Environmental Medicine*, 60, 130-135.

Tsang, P. S., & Vidulich, M. A. (1989). Cognitive Demands of Automation in Aviation. In R. S. Jensen (Ed.), *Aviation Psychology* (pp. 66-95). Brookfield: Gower Technical.

van de Graaff, R. C. (1987). Considerations Concerning the Assessment of Pilot Workload for Complex Task Conditions. *AGARD Conference Proceedings No. 425: The Man-Machine Interface in Tactical Aircraft Design and Combat Automation*, 9-1 - 9-7.

Vidulich, M. A. (1988). The Cognitive Psychology of Subjective Mental Workload. In P. A. Hancock & N. Meshkati (Eds.), *Advances in Psychology: 52. Human Mental Workload* (pp. 219-229). New York: North-Holland.

Vidulich, M. A., & Tsang, P. S. (1986). Techniques of Subjective Workload Assessment: a Comparison of SWAT and the NASA-Bipolar Methods. *Ergonomics*, 29(11), 1385-1398.

Vidulich, M. A., Ward, G. F., & Schueren, J. (1991). Using the Subjective Workload Dominance (SWORD) Technique for Projective Workload Assessment. *Human Factors Society Bulletin*, 33(6), 677-691.

Vidulich, M. A., & Wickens, C. D. (1986). Causes of Dissociation Between Subjective Workload Measures and Performance. *Applied Ergonomics*, 17.4, 291-296.

Vikmanis, M. M. (1987). Advances in Workload Measurement for Cockpit Design Evaluation. *AGARD Conference Proceedings No. 425: The Man-Machine Interface in Tactical Aircraft Design and Combat Automation*, 10-1 - 10-10.

Walker, J. R. (1987). Man-Machine Interface - Operator's Viewpoint. *AGARD Conference Proceedings No. 425: The Man-Machine Interface in Tactical Aircraft Design and Combat Automation*, 1-1 - 1-4.

Wickens, C. D. (1992). Attention, Time-Sharing, and Workload. In C. D. Wickens (Ed.), *Engineering Psychology and Human Performance* (2nd ed., pp. 364-411). New York: Harper Collins.

Wiener, E. L. (1988). Cockpit Automation. In E. L. Wiener & D. C. Nagel (Eds.), *Human Factors in Aviation* (pp. 433-462). New York: Academic Press, Inc.

Wierwille, W. W., & Connor, S. A. (1983). Evaluation of 20 Workload Measures Using a Psychomotor Task in a Moving-Base Aircraft Simulator. *Human Factors Society Bulletin*, 25(1), 1-16.

Wierwille, W. W., & Eggemeier, F. T. (1993). Recommendations for Mental Workload Measurement in a Test and Evaluation Environment. *Human Factors Society Bulletin*, June, 263-281.

Wierwille, W. W., Rahimi, M., & Casali, J. G. (1985). Evaluation of 16 Measures of Mental Workload using a Simulated Flight Task Emphasizing Mediational Activity. *Human Factors Society Bulletin*, 27(5), 489-502.

Wierwille, W. W., & Williges, R. C. (1978). *Survey and Analysis of Operator Workload Assessment Techniques*. Report No. S-78-101 Blacksburg, Virginia: Systemetrics.

Wierwille, W. W., & Williges, B. H. (1980). *An Annotated Bibliography on Operator mental Workload Assessment*. Report No. SY-27R-80 Patuxent River, Maryland: Naval Air Test Center.

Williges, R. C., & Wierwille, W. W. (1979). Behavioral Measures of Aircrew Mental Workload. *Human Factors Society Bulletin*, 21(5), 549-574.

Williges, R. C., Williges, B. H., & Elkerton, J. (1987). Software Interface Design. In G. Salvendy (Ed.), *Handbook of Human Factors* (pp. 1417-1443). New York: John Wiley and Sons.

Williges, R. C., Williges, B. H., & Fainter, R. G. (1988). Software Interfaces for Aviation Systems. In E. L. Wiener & D. C. Nagel (Eds.), *Human Factors in Aviation* (pp. 463-493). New York: Academic Press, Inc.

Winner, R. I., Pennell, J. P., Bertrand, H. E., & Slusarczuk, M. M. G. (1988). *The Role of Concurrent Engineering in Weapons System Acquisition*. IDA Report R-338, Institute for Defense Analysis, Alexandria, Virginia.

Yeh, Y., & Wickens, C. D. (1988). Dissociation of Performance and Subjective Measures of Workload. *Human Factors Society Bulletin*, 30(1), 111-120.

GLOSSARY

ACC	Air Combat Command
AL	Armstrong Laboratory
ANOVA	Analysis of Variance
AO	Attack Operations
ASC	Aeronautical Systems Center
ATC	Automatic Target Cuer
ATR	Automatic Target Recognizer
CE	Concurrent Engineering
CEO	Concurrent Engineering Office
FLIR	Forward Looking Infrared
HCI	Human-Computer Interface
HRGM	High Resolution Ground Map
HSI	Human-System Interface
IDA	Institute for Defense Analysis
IPD	Integrated Product Development
ManTech	Manufacturing Technology Directorate
MCH	Modified Cooper-Harper Scale
NAV	Navigation
NM	Nautical Miles
PAWES	Performance Assessment and Workload Evaluation Systems
ROEs	Rules of Engagement
RSI	Radar Scope Interpretation
SABER	Strategic Avionics Battle-Management Evaluation and Research Lab
SAR	Synthetic Aperture Radar
SME	Subject Matter Expert

SMS	Sensor Management Subsystem
SWAT	Subjective Workload Assessment Technique
SWORD	Subjective Workload Dominance Technique
TLX	NASA Task Load Index
TMD	Theater Missile Defense
VIPER	Visual Image Processing Enhancement and Reconstruction Facility

APPENDIX A

Workload

Performance and Workload

Differentiation:

To utilize experimental data to its fullest, the concepts of performance and workload must be defined and understood as to their application. First, we must differentiate performance from workload. Performance is not always a true and accurate measure of the difficulty of a task. A subject can deal with increased task difficulty by changing the mental or physical effort devoted to the task in such a way that performance remains stable despite the increased difficulty.

As an example, consider the situation where the task is for the subject to successfully cross a gorge or canyon by a variety of means. The first is a standard sturdy bridge that allows passage of motor vehicles. The second is a sturdy foot bridge with hand rails that will accommodate only one pedestrian at a time. The third is a two foot wide bridge of plank with no railings. The subject may cross the gorge successfully for all three (100% performance), but the workload for the third, particularly regarding psychological stress, will differ greatly from the first. As noted in the Handbook of Perception and Human Performance, "To define workload by direct observations on performance, we must conclude that there are no changes in workload even though our common sense and subsequent analysis suggest otherwise" (Gopher & Donchin, 1986).

Performance, as noted in the previous example, may show no variance, yet the workload associated with each task may be different. Additionally, the nature of the workload for each task can be affected by different stressors, e.g., adding a "time factor," altering the "psychological stress factor" by changing the depth of the chasm, or changing the "mental effort" by having the subject repeat a hop-skip-jump movement all the way across the chasm. Moray (1982) noted a study by Tulga in which the operator's performance remained at efficient levels despite a noted increase in the subjective load, as the task's load increased to the point of performance degradation. Moray mentioned an analogy of comparing the operator to a stiff structure that may show little or no change until the point of failure is approached. It would be beneficial to know how near to that

point of failure the operator is, and as noted by Gopher and Donchin (1986) evaluation of mental workload can give that indication.

This concept of workload factors and performance has been shown to be of importance in the automation of aircraft. As technology has progressed and extended the capabilities of the aircraft, the complexity of tasks has increased. The tendency has been to reduce complexity and decrease workload through system automation. This has not always been successful, for automation "redistributes" the workload among the cognitive systems of the operator (Gopher & Donchin, 1986; Kantowitz & Casper, 1988; Matthes, 1987; Reader, 1987; Sexton, 1988; Tsang & Johnson, 1989; Tsang & Vidulich, 1989; Vikmanis, 1987; Walker, 1987; Wiener, 1988).

Automation usually alters the nature of a task by decreasing physical or psychomotor demands and increasing cognitive activities. System functions may be automated, but the operator will still be required to monitor their status. As system performance capabilities grow, the mission capabilities increase, cognitive activities increase, and frequently, there is less time available for their execution.

The effect on operator workload, if any, resulting from a particular system automation, must be compared to the benefits in system performance. Tsang and Johnson (1989) suggested that examination of performance and subjective workload of tasks involving automation would be invaluable in assessing the distribution of cognitive demands and allow identification of any locus of interference between tasks, as well as, intolerable levels of demands. Moray (1982) felt that as the operator became more of a monitor and less of a controller of a system, then subjective workload would be of increasing importance.

Performance, then, is not all that matters in the design of a system. As explained by Vikmanis, (1987) the level of workload matters in the same sense that electronic circuits are designed within their capacity not only to handle required power levels, but also to account for the fact that operating outside the design range results in a higher probability of failure. One must consider what demands a task imposes on the operator's limited resources, since the demands may or may not correspond with performance (Wickens, 1992). This concept of workload is based on the psychological construct that we possess a variety of mental resources which are depleted during the performance of tasks (Vikmanis, 1987). Likewise, if what we are interested in is the performance of the system under different circumstances,

then the element of concern and consideration is mental workload. As explained by Sheridan (1980):

> "... The point is that (as empirical evidence shows) within a certain critical range of mental workload, a small increase in mental workload can result in a large decrement in human and therefore system performance. On the other hand, a system can be performing at almost any level for a given level of operator mental workload, so that the measurement of system performance per se will not predict operator breakdown. Conversely, it is obvious that if its human operator compensates for increasing task difficulty by "trying harder," perfect performance may be found just below the level at which there is a complete breakdown of system performance. This situation cannot be detected by measures of system performance. To predict the breakdown of system performance, one had better measure mental workload."

Definitions

Workload "is a multidimensional, multifaceted, concept that is difficult to define" (Gopher & Donchin, 1986). This concept of workload as a multidimensional construct has been proposed by many authors: Eggemeier, 1981; Johanssen, et al., 1979; Sheridan & Simpson, 1979; Williges & Wierwille, 1979. Much effort has been devoted to define workload and develop suitable measures: Hart, Childress, & Bortolussi., 1981; Johanssen, et al., 1979; Rehmann, Stein, & Rosenberg, 1983; Reid, Shingledecker, & Eggemeier, 1981; Wickens, 1992. This has generated many different measures which are used as workload indicators, but there is no agreement on which of these are the best indicators, nor how these indicators can be used singly or in combination to relate changes in workload (van de Graaff, 1987). Primarily, this is so because no definition of workload can encompass its multidimensional nature (Leplat, 1978; Rehmann, Stein, & Rosenberg, 1983). It follows reason then, that no one measure of a dimension of workload would be complete in defining the workload of a task.

The broad range of activities that workload can encompass, can be and is limited by considering only that which is "mental workload" or those activities requiring primarily mental, rather than physical, effort. Mental workload is seen as a "primitive construct," like happiness, love, and fatigue, which everyone "knows," but yet cannot be defined in precise, operationally useful terms (Jex, 1988).

Papa and Stoliker (1987) viewed workload as a multidimensional construct that was a mixture of each of three interrelated conceptual groups. They noted that most definitions could be placed into one of these groups as follows:

1. those related to the demands of the flight tasks—input load,
2. those associated with the response to those demands—operator effort, and
3. interpretations of workload based on work results or performance.

Sheridan (1980) recommended considering the semantic definition apart from the operational measurement. A single operational definition would not be practical since one must consider the variety of operational definitions used in the measurement of mental workload (Williges & Wierwille, 1979). Thus, the engineer will emphasize definitions based on time and performance, the psychologist will emphasize information processing related to channel capacity and residual attention, and the physiologist will be interested in arousal (Williges & Wierwille, 1979). Unfortunately, most operational definitions observe the task activity as a measure of workload. Again, this measures performance and tends to be insensitive to the variations or conditions that may affect mental workload without affecting the performance (as per the example above).

The problem is to define workload in a measurable sense. Unfortunately, no single activity, signal, measure, or evaluation is considered adequate for defining the whole of workload (Jex, 1988). Lysaght, et al. (1989) describe workload as the elephant being examined by the three blind men of a well-known fable. Each observation about workload is correct, but the conclusion is incorrect or incomplete.

Jex (1988) noted that the main stumbling block to a definition, and hence measurement, was the lack of a proper theoretical framework and models. A discussion of these is beyond the scope of this paper. However, a brief mention of relevant theories and models and how they apply to the definition of mental workload, and hence to subjective measurement, may help in understanding this complex subject.

Stress is often a consequence of high levels of mental workload, especially if sustained. Stress produces changes by its impact on information processing and thereby has effects on performance (Wickens, 1992). Mental workload, as an attribute of the information processing and control systems of the mind, is defined in terms of the limitation on the capacity of these systems. Therefore, the effects of workload on human performance

can be examined only in relation to a model of human information processing. From the Handbook of Perception and Human Performance (Gopher & Donchin, 1986), mental workload is:

> "... the difference between the capacities of the information processing system required for task performance to satisfy performance expectations and the capacity available at any given time. ... Task difficulty is thus manifested by the difference between expected and actual performance. ... The level of expected task performance ... is established by the level of performance of the same task under the least demanding circumstances."

The expected performance may also be derived from knowledge of the relationship between the structure of the task and the nature of the human capacities and skills involved. Alternatively, it may be based on the operator's past performance, or knowledge of the way others perform similar tasks. Frequently, performance expectations are not met and failure in performance is ascribed to increased task difficulty. "Workload" attempts to explain this (Gopher & Donchin, 1986).

In other words, workload accounts for those aspects of the interaction between the operator and a task, such that the task demands exceed the operator's "capacity" to deliver. Workload may thus be viewed as an "attribute of the interaction between a person and a task," and, furthermore, is a "hypothetical construct to summarize the difficulty that a task presents to an operator. ... If the capacity to perform is available, a failure to perform must imply limits to its use. Mental workload is used to explain the mind's inability to deliver" (Gopher & Donchin, 1986). Any measurement of difficulty must consider the interactions between these variables (limited resources and performance), and workload is the label assigned to this interaction.

Workload may then be defined as the relationship between resource supply and task demand. A graphic conceptualization of this relationship shows that in the region to the left of the break point of Figure 2 (adapted from Wickens, 1992), workload is inversely related to the reserve capacity. To the right, workload is inversely related to task performance. According to this concept, workload changes relative to changes in either task resource demands or operator capacity.

The construct of a limited processing capacity is traced to information theory. It is assumed that the capacity of the operator is stable over time and that this capacity or ability

to process information is limited (Lysaght, et al., 1989). If only part of this processing capacity is being used, then there is some capacity not in use. This concept of "spare capacity" implies that if one were "able to measure capacity, then some tasks consume less than full capacity, and leave a residual amount ... that may be measurable" (Gopher & Donchin, 1986). As workload increases, the spare capacity decreases until the point of overload when the information processing demands of the task exceed the total workload capacity.

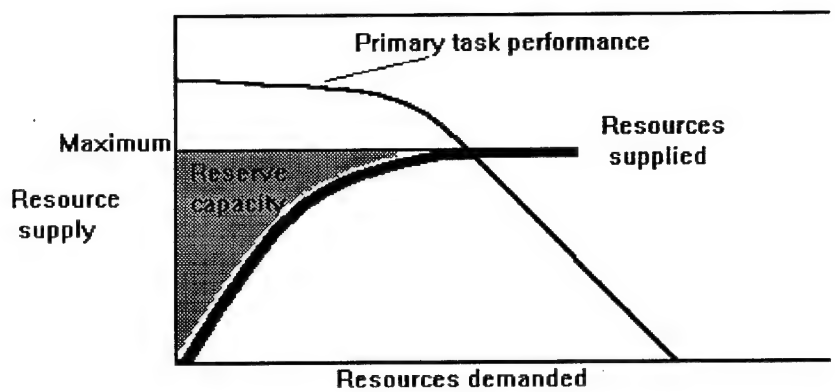

Figure 1. Workload: Resource Supply vs. Task Demand

As explained in the Handbook of Perception and Human Performance, "Different tasks impose different demands on the processor and therefore 'load' it to different extents" (Gopher & Donchin, 1986). Measurements of workload have been developed in an attempt to quantify these "loads" or "capacities," and to predict performance decrements. The implication is that studies are then able to predict which configuration would maximize performance and yet leave the operator with residual capacity to meet unexpected task demands.

Multiple resource theory is another related model based on information processing. It refers to multiple pools of resources or abilities, e.g., verbal or spatial abilities, that may be used singly or in combination in the completion of a task. Workload is thus defined by the utilization of these abilities such that "there will be less competition for the limited resources, and hence less overall workload, when ... not all require the same resource pool ... for processing and controlling..." (Lysaght, et al., 1989).

Some of the problems in measuring this spare capacity are the result of the variety of information processing concepts, e.g., multichannel processing; attention switching between channels; points of conflict or bottlenecks in the information processing channel; and variability in the limits of individual mental workload due to factors such as stress, emotional state, fatigue, and effort (Williges & Wierwille, 1979). Specific measures of workload then "define" workload according to the information processing concept employed, and for that particular operation.

Dissociations

Intercorrelations between performance and workload or workload measures have been described and used to identify and validate measures of workload (Gawron, Schifflet, & Miller, 1989; Sheridan, 1980). Although a high degree of correlation between measures occurs when assessed across tasks of similar structure and varying degrees of difficulty, the correlation may not be present if different tasks are compared (Jex & Clement, 1979; Wickens, 1992).

Dissociations between actual performance and subjective performance (e.g., subjectively rated high, but low/poor actual performance) have been observed and reviewed (Gopher & Donchin, 1986; Wickens, 1992). Yeh and Wickens (1988) noted dissociation between a primary task (performance based) and subjective workload measure. The assumption was that the two factors that were reflected by the subjective measure (i. e., the effort invested to perform a task and the number of concurrent tasks performed) did not necessarily affect performance.

These dissociations are explained in part by the multiple resource theory. Yeh and Wickens (1988) theorized that "performance is determined by (1) amount of resources invested, (2) resource efficiency, and (3) degree of competition for common resources in a multidimensional space described in the multiple resources model." Subjective workload was viewed as multidimensional, increasing when resource investments and working memory demands were increased. Dissociation between performance and subjective workload would occur if more resources were needed for better performance of a resource-limited task, if more working memory was required for time-sharing of tasks (concurrent tasks or display elements), and if performance was easily affected by resource competition, then the subjective measure was found to be better at measuring total resource investment (Yeh & Wickens, 1988).

Returning to Figure 1, if two tasks are in the left sided underload region, more resources will be invested on the more difficult of the two tasks with higher subjective workload, but with no change in performance. Another example is a scenario of three tasks such that full resources are required for all the tasks. Performance will change; however, the subjective workload will be the same since the resources invested (i. e., all) are unchanged. Yet another dissociation can occur if one of two systems compared requires more effort to improve performance, and this effort is "induced" because of monetary incentive, novelty, or better display of information (Wickens, 1992).

Yeh and Wickens (1988) also noted that the number of concurrent tasks had a strong influence on subjective workload, especially regarding time load. If one task, whose difficulty results in poor performance but is quickly performed (low time load), is compared to less difficult, multiple concurrent tasks utilizing different resources that produce good performance, then higher levels of subjective workload (time load) will be noted, reflecting the increased time sharing from the multiple tasks.

Subjective Workload Assessment and Measures

As noted in the Handbook of Perception and Human Performance, a fundamental concept regarding the measurement of workload is that there are characteristics of the subject and task which have an effect on workload and characterize a closed loop, such that the subject's capacities and skills can be assessed by subsequent analysis (e.g., workload assessment techniques or indices) (O'Donnell & Eggemeier, 1986). Subjective workload assessment is one of three major classes of workload assessment techniques. The others are performance-based assessment (e.g., primary and secondary task measures) and physiological workload assessment (Eggemeier, 1981).

Performance-based assessment measures workload by measuring performance changes. These have a restricted range of sensitivity regarding load and workload, as discussed earlier (performance versus workload), and are situation specific (Eggemeier, 1981). Physiological workload has a broad range of measured factors ranging from physical workload affects to mental attentiveness as a measure of workload. These measures are used to measure specific psychological processes in a specific test environment. (Additional information on these measures is outlined by Gawron, Schifflet and Miller in "Measures of In-flight Workload" [Gawron, Schifflet, & Miller, 1989], and by

O'Donnell and Eggemeier in "Workload Assessment Methodology" [O'Donnell & Eggemeier, 1986].) Since the specific interest of this study was to evaluate changes in subjective mental workload, further discussion will center on the concept of subjective mental workload assessment.

As discussed earlier, workload is a multidimensional construct. One advantage of subjective evaluation is that it allows assessment of mental workload to determine what specific area or dimension is most affected by automation and integration of sensors, e.g., time pressure, mental effort (complexity), psychological stress, etc. Subjective measurements have an important function in that they provide information that explores these "dimensions" of the task, and are "almost entirely sufficient in tasks that depend primarily on controlled processes" (O'Donnell & Eggemeier, 1986). According to Sheridan (1980):

> "To predict the breakdown of system performance, one had better measure mental workload. To conform with both everyday semantics and with the investigators own ultimate intent, mental workload should be defined in terms of subjective experience. Subjective scaling is the most direct measure of such subjective experience.... Various other aspects of behavior or biological processes may be measured to serve as predictors of operator and system performance decrement, but they should not be defined to be mental workload."

In his discussion of measuring mental workload, Jex (1988) stated that the definition and measurement of mental workload would center on the subject's activities in that the "...human operator is subjectively aware of his metacontroller activity, and he can introspectively evaluate its 'workload margin' (the excess capacity between the current demands and current metacontroller capability limits)." His definition of this concept was:

> "Mental workload is the operator's evaluation of the attentional load margin (between their motivated capacity and the current task demands) while achieving adequate task performance in a mission-relevant context."

Jex stressed that currently there is no single objective measure. Therefore, the "fundamental measure against which all objective measures must be calibrated is the individual's subjective workload evaluation in each task" (Jex, 1988). The overall consensus and support has been that in any comprehensive workload evaluation, subjective measures must be considered as a central and important element for inclusion (Eggemeier, 1981; Johanssen, et al., 1979; Moray, 1982).

Subjective measures estimate workload through judgments of effort expenditure or similar factors, provided by the operator. These measures have been developed, validated, and used extensively to assess workload, especially in the aviation environment: Bortolussi, Kantowitz, & Hart, 1986; Byers, Bittner, Hill, Zaklad, & Christ, 1988; Casali & Wierwille, 1983; Corwin, 1992; Eggemeier, et al., 1990; Gawron, Schifflet, & Miller, 1989; Hill, et al., 1992; Jex, 1988; Johanssen, et al., 1979; Kantowitz & Casper, 1988; Lysaght, et al., 1989; Moray, 1982; O'Donnell & Eggemeier, 1986; Rehmann, Stein, & Rosenberg, 1983; Reid & Nygren, 1988; Tsang & Johnson, 1989; Tsang & Vidulich, 1989; Vidulich, 1988; Vidulich & Tsang, 1986; Vidulich, Ward, & Schueren, 1991; Vikmanis, 1987; Wickens, 1992; Wierwille & Connor, 1983; Wierwille & Eggemeier, 1993; Wierwille, Rahimi, & Casali, 1985; Williges & Wierwille, 1979. Vikmanis (1987) noted that subjective assessment of workload is particularly effective in simulator evaluations, remarking on the relatively non-intrusive nature of the testing. Wickens (1992) noted that a major advantage of subjective measures is that primary task performance is not disrupted. Furthermore, these measures are globally sensitive techniques, are relatively easy to perform and derive, and have high user acceptance.

Three established subjective measures include the Modified Cooper-Harper Scale (MCH), the NASA Task Load Index (TLX), and the Subjective Workload Assessment Technique (SWAT) (Eggemeier, et al., 1990; Kantowitz & Casper, 1988; Lysaght, et al., 1989; Reid & Nygren, 1988; Reid, Shingledecker, & Eggemeier, 1981; Wickens, 1992). A new measure recently introduced is the Subjective Workload Dominance (SWORD) technique (Vidulich, Ward, & Schueren, 1991).

The MCH is a modification of a sequential subjective rating scale (single dimensional) which was designed to gather information concerning the operator workload and effort involved in performance of a task or group of tasks. Studies in flight simulation have shown it to be of utility in a variety of tasks demanding manipulation including aircraft navigation, communication, and responding to emergency indicators: Casali & Wierwille, 1983; Eggemeier, et al., 1990; Kantowitz & Casper, 1988; Lysaght, et al., 1989; Wickens, 1992; Wierwille & Eggemeier, 1993; Wierwille, Rahimi, & Casali, 1985. The MCH has been found to be a valid and sensitive indicator of overall workload. Its use is recommended when overall mental workload is to be assessed and only in specific experimental situations which do not require subsystem diagnostics (Casali & Wierwille, 1983; Eggemeier, et al., 1990; Gawron, Schifflet, & Miller, 1989).

The NASA TLX, a modification of an earlier NASA Bipolar Scale, involves the use of a series of ratings on six scales (e.g., mental demand, effort, temporal demand, physical demand, perceived performance, and frustration level) to evaluate sub components (dimensions) of operator workload. Workload is defined by this measure in terms of weighted subjective responses (emotional, cognitive, and physical) and weighted evaluation of behaviors (Gawron, Schifflet, & Miller, 1989). It has been used in aviation simulation and actual flight to assess levels of workload associated with different segments of a flight mission (Eggemeier, et al., 1990; Kantowitz & Casper, 1988; Lysaght, et al., 1989; Wickens, 1992; Wierwille & Eggemeier, 1993). There have been some problems and deficiencies related to the calculation of the correlations and skewed distributions of subjective responses in a study reviewed by Gawron, et al. (1989). However, TLX can be used as a sensitive indicator of overall workload, and may be more useful in an operational environment where time to complete the assessment may be a concern. Interestingly, Gawron notes that four TLX dimensions shown to be of most importance by subjects were "in parallel" with the three SWAT dimensions. Nygren (1991) felt that the dimensional weighting procedure of TLX was ineffective and should not be used, but was capable in applied settings such as systems development.

SWAT is also multidimensional, in that it rates three components or dimensions of subjective workload: time load, mental effort load, and stress load. It has been used extensively in a variety of aviation environments, including simulators, (Corwin, 1992; Eggemeier, et al., 1990; Lysaght, et al., 1989; Reid & Nygren, 1988; Reid, Shingledecker, & Eggemeier, 1981; Vikmanis, 1987; Wickens, 1992; Wierwille & Eggemeier, 1993) and specifically, in recent studies of single sensor integration and automation (Kuperman, 1992; Papa & Stoliker, 1987). SWAT and similar measures were developed with the objective of providing a tool to evaluate these types of technological alternatives in ground-based simulators and flight tests (Vikmanis, 1987). SWAT is a valid, sensitive and unobtrusive measure (Eggemeier, et al., 1990; Gawron, Schifflet, & Miller, 1989; Lysaght, et al., 1989; Reid & Nygren, 1988; Reid, Shingledecker, & Eggemeier, 1981). A more detailed explanation of SWAT is presented in the following section.

These three techniques of measurement differ by the number of dimensions or subscales evaluated, the number of ratings for each dimension (which may provide diagnostic information on the source of workload), and the algorithm for calculating the workload. They evaluate each task separately and are designed to be used after completion of the task

(innately non-intrusive). They are all considered to be globally sensitive measures with high user acceptance.

SWORD is a new arrival to subjective evaluation techniques. It uses a series of relative judgments to compare the workload of different task situations. This is in contrast to the other techniques that use absolute estimation, e.g., the operator assigns a value from a workload scale to a task which is not in relative comparison to any other task. Because of its recent development, its full utility in various settings and environments has yet to be evaluated.

However, subjective measures are not without potential problems. One potential drawback is immediacy. Once the task is completed, the assessments must be given in a timely manner in order to accurately reflect and measure the workload (Vidulich & Tsang, 1986).

As discussed earlier, there can be some dissociation between subjective workload measures and performance (Adelman & Donnell, 1988; Lysaght, et al., 1989; Vidulich & Wickens, 1986; Wickens, 1992; Yeh & Wickens, 1988). Vidulich and Wickens (1986) found that dissociations between workload and performance were related to automaticity, presentation rate, and motivation.

In a multi-task study automaticity dissociation might reduce the usefulness of subjective techniques to detect the influence of one sub-task. Thus, subjective techniques are best used to evaluate side-tasks in a single-task environment. Apparently a dissociation occurs when there is a rate-change manipulation. In this case subjects tend to rate the slower conditions as easier, even if performance is less efficient. The changes, seen by differing levels of "motivation," are of concern when one of the tasks under evaluation may be perceived as more interesting or novel than the other(s). Here the subjects may increase the available resources, thus aiding performance in the novel task and may therefore experience less fatigue (Moray, 1982).

Evaluations of subjective techniques have demonstrated their utility in both flight simulation and the actual flight environment (Eggemeier, et al., 1990; Kantowitz & Casper, 1988; Lysaght, et al., 1989; Papa & Stoliker, 1987; Vikmanis, 1987; Wickens, 1992). Comparative evaluations have not shown significant advantages of one over the other (Bortolussi, Kantowitz, & Hart, 1986; Byers, Bittner, Hill, Zaklad, & Christ, 1988; Casali

& Wierwille, 1983; Eggemeier, et al., 1990; Hill, et al., 1992; Kantowitz & Casper, 1988; Lysaght, et al., 1989; Moray, 1982; Nygren, 1991; Tsang & Johnson, 1989; Vidulich & Tsang, 1986; Wickens, 1992; Wierwille & Connor, 1983; Wierwille & Eggemeier, 1993; Wierwille, Rahimi, & Casali, 1985; Wierwille & Williges, 1978; Wierwille & Williges, 1980). Basically, much depends on what is being evaluated and what is to be expected out of that evaluation. SWAT, however, has had a long and successful history of application in United States Air Force workload assessment, and has been recommended by the PAWES study (Performance Assessment and Workload Evaluation Systems) as the technique to be used for subjective assessment (Eggemeier, et al., 1990). Furthermore, the extensive use of SWAT in various settings may allow some retrospective comparison between studies. Additionally, dimensional weighting is better evaluated by SWAT and, in this instance, is preferred over TLX.

In a paper on recommendations for mental workload measurement, Wierwille and Eggemeier (1993) stressed that part of the problem is the multitude of techniques and experiments, as well as many conceptual issues that are complex and still unclarified. The various factors, as mentioned above, that may affect subjective techniques can be minimized by selecting subjects that have similar experience and have performed under an equivalent range of system conditions. The consensus was that in a simulated flight environment, there was no marked or consistent sensitivity difference between the techniques. Nygren (1991) indicated that SWAT seemed to be better at identifying factors such as cognitive mechanisms affecting mental workload judgments, while TLX was better in settings of low levels of workload when compared to SWAT. However, TLX demonstrated consistently higher loadings than SWAT or MCH. Compared to MCH, TLX and SWAT have the advantage of being multidimensional and can therefore provide additional information as to the source of workload.

Yeh and Wickens (1988) also wrote guidelines in the use of subjective workload techniques based on the insensitivity of these measures in tasks of high cognitive load and common resource competition, combined with their sensitivity to invested resources and time-sharing. When task demand is in the underload region, subjective measures accurately reflect reserve capacity. But, when the cognitive load exceeds working memory capacity, subjective measures are less sensitive and primary task performance measures are preferred. In the case of resource limited tasks, these researchers felt that the determining factors regarding subjective measure were the total amount of invested resources and, especially, the demand on working memory. Therefore, subjective measures would show bias in

situations that required time-sharing between concurrent tasks and would not be sensitive to changes in performance that were the result of common resource competition (Yeh & Wickens, 1988).

This study was designed to eliminate any potential bias as noted by Yeh and Wickens (e.g., no time-sharing between tasks, working memory capacity is not exceeded and task demand is in the underload region to reflect reserve capacity). With the intention of evaluation of the various factors that affect workload by changes in the sensor format, and in accordance with Nygren, et al. (1991) and Eggemeier, et al. (1990), SWAT was selected as the technique that would best evaluate subjective workload in this particular instance.

Subjective Workload Assessment Technique

As stated earlier, the sensitivity of this technique has been well documented. It has demonstrated the ability to differentiate between a variety of task demands over a range of information processing functions, and in both the simulated and actual environments. Due to its excellent sensitivity, it represents a globally sensitive measure of operator mental workload that can evaluate the various dimensions of workload. Although it is not considered to be diagnostic in discriminating levels of load on individual information processing functions, it is diagnostic regarding the dimensions of workload measured. Also, it is able to demonstrate sensitivity to variations in task demand in all of the generic functions of the aviator (Eggemeier, et al., 1990; Papa & Stoliker, 1987).

Because of SWAT's multidimensional nature, it provides information concerning the sources of levels of load associated with each task, in that the three subscales or dimensions are differentially affected by different variables (Eggemeier, et al., 1990). Analysis can thus provide some information regarding the factor (time, effort, or stress) that most influenced the workload rating, and can be used to provide some diagnostic information concerning the source of workload. Because the SWAT task evaluation (event scoring) is given immediately after completion of the task to be evaluated, it is a relatively non-intrusive technique. Additionally, it requires few materials to implement, and is relatively well accepted by subjects.

SWAT relates to the relative perceived workload in three dimensions:

1. Time load: refers to the fraction of time the subject is occupied by task work, or the amount of time one has available to perform a task(s). When the load is low, there is adequate time to complete all the mental work with time to spare. As time load increases, spare time decreases and tasks may overlap and interrupt each other, e.g., performing more than one task or different aspects of the same task. As time load becomes more critical, even more aspects of task performance occur simultaneously and interruptions are more frequent. The rate that events occur or the speed of the system may also be reflected by time load.

2. Mental effort load: refers to the amount of attention or mental effort required without regard to the number of tasks to be performed or time limitations. When the load is low, the concentration and attention required for the task is low and performance may be automatic. As mental effort load increases due to complexity of the task or the amount of information requiring integration, then the amount of concentration and attention needed will increase until it is utilized totally at high loads. Problem solving, calculations, decision making and storing and recalling information will be reflected by mental load.

3. Psychological stress load: refers to the contribution from stressors that produce anxiety, frustration, or confusion during task performance. As the level of stress load increases, producing an increase in anxiety, frustration, or confusion, greater effort, regarding concentration and determination, is required of the test subject to maintain situational control. At low stress levels the subject is relaxed with total attention to the task. As the level or stress increases, distraction occurs. Factors reflecting this load include motivation, fatigue, fear, skill level, temperature, noise, vibration, and comfort. At high levels these can impact task performance. In the case of SWAT, these factors are considered to be at low levels, just enough to be an irritant, such that the subject must use resources to prevent any interference with task performance.

Each of the three dimensions is further divided into three levels to relate to high, medium, and low loading. These have been defined as follows:

I. **Time load**

1. Often have spare time. Interruptions or overlap among activities occur infrequently or not at all.

2. Occasionally have spare time. Interruptions or overlap among activities occur frequently.

3. Almost never have spare time. Interruptions or overlap among activities are very frequent or occur all the time.

II. **Mental effort**

1. Very little conscious mental effort or concentration required. Activity is almost automatic requiring little or no attention.

2. Moderate conscious mental effort or concentration required. Complexity of activity is moderately high due to uncertainty, unpredictability or unfamiliarity. Considerable attention required.

3. Extensive mental effort and concentration are necessary. Very complex activity requiring total attention.

III. **Psychological stress**

1. Little confusion, frustration or anxiety exists and can be easily accommodated.

2. Moderate stress due to confusion, frustration, or anxiety. Noticeably adds to workload. Significant compensation is required to maintain adequate performance.

3. High to very intense stress due to confusion, frustration, or anxiety. High to extreme determination and self-control required.

The SWAT cards are labeled according to these definitions with 27 cards listing all the possible combinations of the three levels of the three subscales or dimensions. The subjects rank order the cards (workload descriptors) according to their perception of workload for a particular task, called the "card sort." This is the key step in the first phase of SWAT, called the (individual) **scale development** phase of SWAT. These individual card sorts from each subject are combined through conjoint measurement to test the ordered data, and to determine the combination rules used by each of the individuals in their card sort.

A SWAT computer program is used to analyze each subject's data, thus allowing determination of the individual subject's prototype (which dimension contributes most to perceived workload). A group scaling may also be performed in which all the data are averaged together and a single group scale is derived by conjoint analysis. (A more detailed explanation on the development and validation of SWAT and its application of conjoint analysis is provided in a paper by Reid, Shingledecker and Eggemeier [1981], in a paper by Nygren, [1991] and in "...SWAT: a user's guide" by Reid, Potter and Bressler. [1989].)

The second phase of SWAT in which the subject evaluates each task, rating each dimension according to a three-point rating scale, is referred to as event scoring. These

event scores are then translated into individual components related to time, effort, or stress load dimensions, and an overall workload score.

An additive model produces the combined workload score, which incorporates the three load dimensions on an interval scale of zero to 100. This aspect of SWAT is unique among the subjective measures, in that it is based on a tested model of how judgments of mental workload are done. The model, as explained by Nygren, "assumes that perceived mental workload is equivalent to an additive combination of the effects of three psychologically relevant dimensions, the validity of which is testable using conjoint measurement..." (Nygren, 1991). Because of this basis in additive conjoint measurement, SWAT requires only that the raters think of the dimensions as independent of each other, and does not require that the dimensions be statistically independent or uncorrelated in the actual tasking. Also, because of its conditional monotonicity, higher values on each dimension tend to predict higher levels with regard to performance independent of the values of the other dimensions (Nygren, 1991).

Reid and Colle (1988) reviewed previous studies and found that there was a range of SWAT scores that were predictive of operator overload. Nevertheless, researchers have found that a SWAT score of 40 plus or minus 10 (range 30-50) could be used to indicate potential performance difficulties due to workload saturation and overload (Reid, Potter, & Bressler, 1989).

The final analysis may then show not only how workload was affected in terms of degree or level perceived, but which dimensions of workload were involved in this change, as well. As noted previously, the outcome may be affected by certain biases common to subjective workload techniques. In addition to the problems of immediacy and dissociation, there may in some cases be correlations between the three scales. One explanation has been that at a low level of cognitive load, primarily mental effort was affected, but, as the task load increased, changes were also noted in the areas of time and psychological stress (Tsang & Vidulich, 1989). Gawron, et al. (1989) mentioned a study by Boyd which also suggested that these workload dimensions were not independent in certain circumstances.

APPENDIX B

Rating Scale SME Comments

This appendix includes the transcribed SME comments volunteered during the administration of the rating scale questionnaires. It also presents the SME responses to a background and experience questionnaire. (It should be noted that, although data from only 14 SMEs were included in the analyses, a 15th SME [SME 05] completed the debriefing comments portion of the questionnaire.)

SENSOR INTEGRATION QUESTIONNAIRE

RUN 1 : SAR/HRGM

1. If you were using real avionics in an actual aircraft, how well would you be able to DETECT (target position)?

SUBJECT	COMMENT
01	Never did any target detection, location, identification (arrow to include all).
02	No Comment.
03	No Comment.
04	No Comment.
05	Assuming the spots were targets.
06	No Comment.
07	Good contrast between ground return and potential target.
08	You can detect a radar return, but not a type of target.
09	No Comment.
10	No Comment.
11	No Comment.
12	No Comment.
13	No Comment.
14	No Comment
15	Easy to see spots.

2. If you were using real avionics in an actual aircraft, how well would you be able to LOCATE (handoff target position/location to a weapon)?

SUBJECT	COMMENT
01	As above.
02	No Comment.
03	No Comment.
04	Only gravity weapon experience.
05	No Comment.
06	No Comment.
07	Very easy to use TH system to move cursor over potential target.
08	Very broad coordinate area.
09	No Comment.
10	No Comment.
11	No Comment.
12	No Comment.
13	No Comment.
14	No Comment
15	Easy to position cursor.

3. If you were using real avionics in an actual aircraft, how well would you be able to IDENTIFY (target)?

SUBJECT	COMMENT
01	As above.
02	No Comment.
03	Two targets were easy to identify. The third was not "linear" nor every bright. Could have been a tree. Chose not to launch.
04	10' SAR not good at ID target.
05	Nothing to compare relative size. I designated spots-not readily identifiable targets.
06	No Comment.
07	No much resolution to tell if target was a TEL, decoy, or other ground vehicle, great contrast, but no solid ID capability.
08	No shape or physical characteristics distinguishable.
09	No Comment.
10	No Comment.
11	No Comment.
12	Hard to determine what the target actually is.
13	No Comment.
14	No Comment
15	I can see bright spots. but can't tell what they are.

4. WORKLOAD: How did this avionics capability affect your workload?

SUBJECT	COMMENT
01	Again, nothing to really compare to.
02	Color cues for auto map cueing desirable.
03	No Comment.
04	No Comment.
05	No baseline to compare.
06	No Comment.
07	Very easy to use cut and dry methods/procedures.
08	No Comment.
09	No Comment.
10	No Comment.
11	No Comment.
12	Not time consuming or overbearing.
13	No Comment.
14	No Comment
15	No Comment.

5. TRAINING: What effect would this avionics capability have on training requirements?

SUBJECT	COMMENT
01	No Comment.
02	SAR training/deployment tactics.
03	No Comment.
04	No Comment.
05	Unknown.
06	No Comment.
07	I'm not sure some increase in training would be necessary to inform the user of the use and implementation of this technology.
08	Using the sensor manager takes minimal training.
09	No Comment.
10	No Comment.
11	No Comment.
12	I would want to initially know more functional attributes of the screens, but after that was understood, the actual use of the screens doesn't require extensive training.
13	No Comment.
14	No Comment
15	No Comment.

6. CONFIDENCE: How did this affect your mission confidence?

SUBJECT	COMMENT
01	No Comment.
02	No Comment.
03	Compared to NAV update.
04	No Comment.
05	No baseline.
06	No Comment.
07	Seemed to make target selection very straight forward. No interpretation needed.
08	Slightly enhanced perception that a target of some sort is there.
09	No Comment.
10	No Comment.
11	No Comment.
12	No Comment.
13	No Comment.
14	No Comment
15	No Comment.

7. SITUATIONAL AWARENESS: How did this affect your SA?

SUBJECT	COMMENT
01	No Comment.
02	No Comment.
03	No Comment.
04	No Comment.
05	No Comment.
06	No Comment.
07	Large amount of information presented, some what "cluttered" display, not in format of presentation, but in amount of info present.
08	No Comment.
09	No Comment.
10	No Comment.
11	No Comment.
12	No Comment.
13	No Comment.
14	No Comment
15	No Comment.

8. PACING: How did this affect your mission pacing?

SUBJECT	COMMENT
01	This would probably be hard to estimate since I am doing so little except the target acquisition.
02	No Comment.
03	No Comment.
04	No Comment.
05	No baseline.
06	No Comment.
07	No Comment.
08	No Comment.
09	No Comment.
10	No Comment.
11	No Comment.
12	No Comment.
13	No Comment.
14	No Comment
15	No Comment.

9. FALSE ALARMS: How did this affect false alarm occurrence?

SUBJECT	COMMENT
01	No Comment.
02	No Comment.
03	But still not completely sure.
04	No Comment.
05	Unknown- I would have designated false targets.
06	No Comment.
07	No Comment.
08	Still cannot distinguish type of images - no recognition or identification.
09	No Comment.
10	No Comment.
11	No Comment.
12	No Comment.
13	No Comment.
14	No Comment
15	No Comment.

10. SENSOR MANAGER/ATC: How did this affect the mission?

SUBJECT	COMMENT
01	No Comment.
02	No Comment.
03	No Comment.
04	No Comment.
05	Don't know. I presumed it reduces the effort to locate target. No baseline.
06	No Comment.
07	Made mission profile easy to execute.
08	No Comment.
09	No Comment.
10	No Comment.
11	No Comment.
12	No Comment.
13	No Comment.
14	No Comment
15	No Comment.

11. IN-THE-LOOP-CONTROL: How did this affect your "feeling" of in-the-loop control?

SUBJECT	COMMENT
01	No Comment.
02	No Comment.
03	No Comment.
04	No Comment.
05	No Comment.
06	No Comment.
07	Felt in control of the event the whole time.
08	Targeting is an active function; but once again, determining if it is an actual military target is difficult to impossible from the imagery.
09	No Comment.
10	No Comment.
11	No Comment.
12	No Comment. ** No previous experience with target detection to compare this run to. **
13	No Comment.
14	No Comment
15	No Comment.

RUN 2 : SAR/HRGM + SPOT

1. If you were using real avionics in an actual aircraft, how well would you be able to DETECT (target position)?

SUBJECT	COMMENT
01	No Comment.
02	No Comment.
03	Spotlight really helps for confirmation.
04	No Comment.
05	Better than run #1.
06	No Comment.
07	Again good ground/target controls.
08	SPOT imagery defines shape of possible target.
09	No Comment.
10	No Comment.
11	No Comment.
12	No Comment.
13	No Comment.
14	No Comment
15	No Comment.

2. If you were using real avionics in an actual aircraft, how well would you be able to LOCATE (handoff target position/location to a weapon)?

SUBJECT	COMMENT
01	No Comment.
02	No Comment.
03	No Comment.
04	Gravity only.
05	No Comment.
06	No Comment.
07	No Comment.
08	Much better capability to define position for weapon than with HRGM.
09	No Comment.
10	No Comment.
11	No Comment.
12	No Comment.
13	No Comment.
14	No Comment
15	No Comment.

3. If you were using real avionics in an actual aircraft, how well would you be able to IDENTIFY (target)?

SUBJECT	COMMENT
01	No Comment.
02	Spot image helped identify a false alarm for me.
03	No Comment.
04	Improved ability to locate target position i.e. near tree line, road, etc...
05	Nothing to compare relative size. Things designated were no longer spots- they were short lines in the form of TELs although I questioned positioning of them given respect to surrounding cover.
06	No Comment.
07	SPOT function gives much better view of potential target.
08	Cannot type the target - put it into it's class.
09	No Comment.
10	No Comment.
11	No Comment.
12	2.5' enhancement allows for better identification of target (obviously huh?).
13	No Comment.
14	No Comment
15	Able to get a better "view" of target.

4. WORKLOAD: How did this avionics capability affect your workload?

SUBJECT	COMMENT
01	No Comment.
02	No Comment.
03	Less uncertainty - less mental effort to decide --just spotlight it -- fast & easy.
04	No Comment.
05	No baseline.
06	No Comment.
07	Slight increase in workload necessary to scan SPOT image.
08	Do not have to search entire area to find area of interest.
09	No Comment.
10	No Comment.
11	No Comment.
12	No Comment.
13	No Comment.
14	No Comment.
15	No Comment.

5. TRAINING: What effect would this avionics capability have on training requirements?

SUBJECT	COMMENT
01	No Comment.
02	SAR training.
03	No Comment.
04	No Comment.
05	Unknown.
06	No Comment.
07	Shouldn't be too difficult to master, very similar to HRGM images.
08	Same training as for HRGM.
09	No Comment.
10	No Comment.
11	No Comment.
12	No Comment.
13	No Comment.
14	No Comment.
15	No Comment.

6. CONFIDENCE: How did this affect your mission confidence?

SUBJECT	COMMENT
01	No Comment.
02	Spot made it absolute.
03	No Comment.
04	No Comment.
05	No Comment.
06	No Comment.
07	Gave a more precise picture of potential target (using SPOT function).
08	Much more confident in distinguishing targets - closer to recognition than ID.
09	No Comment.
10	No Comment.
11	No Comment.
12	No Comment.
13	No Comment.
14	No Comment.
15	No Comment.

7. SITUATIONAL AWARENESS: How did this affect your SA?

SUBJECT	COMMENT
01	No Comment.
02	No Comment.
03	No Comment.
04	No Comment.
05	No Comment.
06	No Comment.
07	No Comment.
08	More aware of target environment -- however, may detract from actual flying due to short/small time looking at SPOT imagery.
09	No Comment.
10	No Comment.
11	No Comment.
12	No Comment.
13	No Comment.
14	No Comment.
15	No Comment.

8. PACING: How did this affect your mission pacing?

SUBJECT	COMMENT
01	No Comment.
02	No Comment.
03	No Comment.
04	No Comment.
05	No Comment.
06	No Comment.
07	Concerned a bit on how much time additional spotting would take but seemed to not use up a lot of additional time.
08	No Comment.
09	No Comment.
10	No Comment.
11	No Comment.
12	No Comment.
13	No Comment.
14	No Comment.
15	No Comment.

9. FALSE ALARMS: How did this affect false alarm occurrence?

SUBJECT	COMMENT
01	No Comment.
02	No Comment.
03	No Comment.
04	No Comment.
05	No Comment.
06	No Comment.
07	No Comment.
08	Better imagery to see potential targets.
09	Reduces the occurrence of false alarms.
10	No Comment.
11	No Comment.
12	No Comment.
13	No Comment.
14	No Comment.
15	No Comment.

10. SENSOR MANAGER/ATC: How did this affect the mission?

SUBJECT	COMMENT
01	No Comment.
02	No Comment.
03	No Comment.
04	No Comment.
05	No Comment.
06	Noticed one target was not registered between HRGM and SPOT displays - caused momentary confusion.
07	No Comment.
08	No Comment.
09	No Comment.
10	No Comment.
11	No Comment.
12	No Comment.
13	No Comment.
14	No Comment.
15	No Comment.

11. IN-THE-LOOP-CONTROL: How did this affect your "feeling of in-the-loop control?

SUBJECT	COMMENT
01	No Comment.
02	No Comment.
03	Image gives me what I need to see -- I make the decision.
04	No Comment.
05	No Comment.
06	No Comment.
07	No Comment.
08	No Comment.
09	No Comment.
10	No Comment.
11	No Comment.
12	No Comment.
13	No Comment.
14	No Comment.
15	No Comment.

RUN 3 : SAR/HRGM + SPOT + ATC

1. If you were using real avionics in an actual aircraft, how well would you be able to DETECT (target position)?

SUBJECT	COMMENT
01	No Comment.
02	No Comment.
03	Helps know where to start looking.
04	No Comment.
05	Better than run # 1.
06	No Comment.
07	Good target/background contrast.
08	With aid of SPOT imagery detection is easier.
09	No Comment.
10	No Comment.
11	No Comment.
12	No Comment.
13	No Comment.
14	No Comment.
15	No Comment.

2. If you were using real avionics in an actual aircraft, how well would you be able to LOCATE (handoff target position/location to a weapon)?

SUBJECT	COMMENT
01	No Comment.
02	No Comment.
03	No Comment.
04	No Comment.
05	No Comment.
06	No Comment.
07	No Comment.
08	No Comment.
09	No Comment.
10	No Comment.
11	No Comment.
12	No Comment.
13	No Comment.
14	No Comment.
15	No Comment.

3. If you were using real avionics in an actual aircraft, how well would you be able to IDENTIFY (target)?

SUBJECT	COMMENT
01	No Comment.
02	No Comment.
03	No Comment.
04	No Comment.
05	Still nothing of relative size to judge from. Same comments as run #2: Nothing to compare relative size. Things designated were no longer spots- they were short lines in the form of TELs although I questioned positioning of them given respect to surrounding cover.
06	No Comment.
07	Seemed to automatically identify targets fine.
08	No Comment.
09	No Comment.
10	No Comment.
11	No Comment.
12	No Comment.
13	No Comment.
14	No Comment.
15	No Comment.

4. WORKLOAD: How did this avionics capability affect your workload?

SUBJECT	COMMENT
01	No Comment.
02	No Comment.
03	From baseline (no diff from run 2); Still scanning to see if there's a possible target that the ATC didn't catch.
04	Learning curve. In long run it should decrease.
05	Would greatly reduce workload if only the circles contained targets. If time was serious limitation, circles would greatly help. In this case I could have searched whole area manually, given the time I had.
06	Confusion of control sequence - change from previous - caused display mismatch.
07	Increased workload, especially when operator error occurred. Was unsure of where/what error was. This probably could be reduced with further training.
08	Sequentially viewing possible targets takes a little longer.
09	No Comment.
10	No Comment.
11	No Comment.
12	No Comment.
13	No Comment.
14	No Comment.
15	Sequenced spot image made it harder than with my control over spot image.

5. TRAINING: What effect would this avionics capability have on training requirement?

SUBJECT	COMMENT
01	No Comment.
02	You'll have crew members who will want to know the math behind how cuer determines target.
03	Some learning; not much. Might want to know how to disable the ATC part, if desired.
04	No Comment.
05	Unknown.
06	Mainly from negative transfer from previous configuration.
07	If ATC were used, less training required for manual targeting ID.
08	No Comment.
09	No Comment.
10	No Comment.
11	No Comment.
12	No Comment.
13	No Comment.
14	No Comment.
15	No Comment.

6. CONFIDENCE: How did this affect your mission confidence?

SUBJECT	COMMENT
01	No Comment.
02	No Comment.
03	No Comment.
04	No Comment.
05	No Comment.
06	No Comment.
07	Again, due mostly to operator error, "lost train of thought" when things didn't go as expected.
08	No Comment.
09	No Comment.
10	No Comment.
11	No Comment.
12	No Comment.
13	No Comment.
14	No Comment.
15	No Comment.

7. SITUATIONAL AWARENESS: How did this affect your SA?

SUBJECT	COMMENT
01	No Comment.
02	No Comment.
03	No Comment.
04	No Comment.
05	No Comment.
06	No Comment.
07	More confusion than in previous runs, again for the above mentioned reason.
08	No Comment.
09	No Comment.
10	No Comment.
11	No Comment.
12	No Comment.
13	No Comment.
14	No Comment.
15	No Comment.

8. PACING: How did this affect your mission pacing?

SUBJECT	COMMENT
01	No Comment.
02	No Comment.
03	No Comment.
04	No Comment.
05	No Comment.
06	No Comment.
07	Still seems to be plenty of time to accomplish tasks.
08	Sequential search slows down targeting.
09	No Comment.
10	No Comment.
11	No Comment.
12	No Comment.
13	No Comment.
14	No Comment.
15	I was unsure of sequence.

9. FALSE ALARMS: How did this affect false alarm occurrence?

SUBJECT	COMMENT
01	No Comment.
02	No Comment.
03	No Comment.
04	No Comment.
05	No Comment.
06	Drawn to false target I would have ignored.
07	No Comment.
08	Can verify no target exists in a scene with SPOT imagery.
09	Reduced the occurrence of false alarms.
10	No Comment.
11	No Comment.
12	No Comment.
13	No Comment.
14	No Comment.
15	No Comment.

10. SENSOR MANAGER/ATC: How did this affect the mission?

SUBJECT	COMMENT
01	No Comment.
02	No Comment.
03	No Comment.
04	No Comment.
05	No Comment.
06	No Comment.
07	Was not able to select all 3 targets due to operator error.
08	No Comment.
09	No Comment.
10	No Comment.
11	No Comment.
12	No Comment.
13	No Comment.
14	No Comment.
15	No Comment.

11. IN-THE-LOOP-CONTROL: How did this affect your "feeling" of in-the-loop control?

SUBJECT	COMMENT
01	No Comment.
02	No Comment.
03	If I can still look at possible targets that the ATC does not identify. (If I can only look at ATC target circles, and can't call up the 2.5' on other targets, then the MIL is degraded.).
04	No Comment.
05	No Comment.
06	Letting the machine do more, I'm doing less.
07	Again, operator recognized an error occurred, but didn't have strong idea on how to correct problem. I did not feel that ATC by itself deprived user of any control authority.
08	No Comment.
09	No Comment.
10	No Comment.
11	No Comment.
12	Would like the real time capability of returning to a target area. Don't like computer limiting my search area.
13	No Comment.
14	No Comment.
15	No Comment.

RUN 4 : SAR/HRGM + SPOT + ATC + FLIR

1. If you were using real avionics in an actual aircraft, how well would you be able to DETECT (target position)?

SUBJECT	COMMENT
01	No Comment.
02	No Comment.
03	No Comment.
04	No Comment.
05	Better than run #1, 2, or 3.
06	No Comment.
07	Same as other runs, no addition detection ability.
08	No Comment.
09	No Comment.
10	No Comment.
11	No Comment.
12	No Comment.
13	No Comment.
14	No Comment.
15	No Comment.

2. If you were using real avionics in an actual aircraft, how well would you be able to LOCATE (handoff target position/location to a weapon)?

SUBJECT	COMMENT
01	No Comment.
02	No Comment.
03	No Comment.
04	No Comment.
05	No Comment.
06	No Comment.
07	Automatic locate function (ATC) was helpful.
08	No Comment.
09	No Comment.
10	No Comment.
11	No Comment.
12	No Comment.
13	No Comment.
14	No Comment.
15	No Comment.

3. If you were using real avionics in an actual aircraft, how well would you be able to IDENTIFY (target)?

SUBJECT	COMMENT
01	No Comment.
02	No Comment.
03	No Comment.
04	No Comment.
05	Much better. Still lack relative size. Different imagery enhances confidence.
06	No Comment.
07	FLIR presented good images. Definite determination as to the false target in Run 4.
08	No Comment.
09	No Comment.
10	No Comment.
11	No Comment.
12	FLIR adds element of uncertainty in some situations (target #2).
13	FLIR imagery so poor it added noise to the decision process.
14	No Comment.
15	The added input of the FLIR was very confusing for target #2. This made it more difficult to tell if it was a target.

4. WORKLOAD: How did this avionics capability affect your workload?

SUBJECT	COMMENT
01	No Comment.
02	No Comment.
03	About equal with baseline. A little more time consuming than runs 2 or 3.
04	No Comment.
05	Increase workload over #3 run.
06	Extra step increases workload.
07	No extra effort seemed required to use FLIR.
08	Slightly increased in calibrating eye to range change then imagery change.
09	No Comment.
10	No Comment.
11	No Comment.
12	Same as #3.
13	No Comment.
14	No Comment.
15	No Comment.

5. TRAINING: What effect would this avionics capability have on training requirements?

SUBJECT	COMMENT
01	No Comment.
02	No Comment.
03	No Comment.
04	No Comment.
05	Unknown.
06	No Comment.
07	Necessitate training on FLIR image interpretation.
08	No Comment.
09	No Comment.
10	No Comment.
11	No Comment.
12	No Comment.
13	No Comment.
14	No Comment.
15	No Comment.

6. CONFIDENCE: How did this affect your mission confidence?

SUBJECT	COMMENT
01	No Comment.
02	No Comment.
03	No Comment.
04	No Comment.
05	Better capability than run # 1, 2, or 3.
06	Much more confident of target.
07	Presented further information to user to confirm target.
08	Much more confident in being able to locate and detect. Identification is a bit more time consuming, but weapons could be launched with greater confidence.
09	No Comment.
10	No Comment.
11	No Comment.
12	No Comment.
13	Again FLIR confused issue because of poor quality.
14	No Comment.
15	For 2 of the targets (sectors 1 & 3) I felt it enhanced confidence, but the overall effect was negative since I was so unsure of target #2.

7. SITUATIONAL AWARENESS: How did this affect your SA?

SUBJECT	COMMENT
01	No Comment.
02	No Comment.
03	No Comment.
04	No Comment.
05	No Comment.
06	No Comment.
07	No Comment.
08	FLIR imagery did not greatly enhance location from SPOT imagery as the scaling is different. SA is affected by trying to gauge position off different sensors.
09	No Comment.
10	No Comment.
11	No Comment.
12	No Comment.
13	No Comment.
14	No Comment.
15	I didn't trust the FLIR, but when I saw good confirmation, I felt SA was enhanced.

8. PACING: How did this affect your mission pacing?

SUBJECT	COMMENT
01	No Comment.
02	Adding additional sensor compressed time.
03	No Comment.
04	No Comment.
05	No Comment.
06	No Comment.
07	Again, no extra time needed to use FLIR.
08	FLIR imagery takes a bit longer to process after looking at HRGM & SPOT imagery.
09	No Comment.
10	No Comment.
11	No Comment.
12	No Comment.
13	FLIR took extra time.
14	No Comment.
15	No comment.

9. FALSE ALARMS: How did this affect false alarm occurrence?

SUBJECT	COMMENT
01	No Comment.
02	No Comment.
03	No Comment.
04	No Comment.
05	No Comment.
06	No Comment.
07	Gave ability to detect false automatic targeting.
08	Can detect as well in SPOT.
09	Reduced occurrence of false alarm.
10	No Comment.
11	No Comment.
12	No Comment.
13	No Comment.
14	No Comment.
15	No comment.

10. SENSOR MANAGER/ATC: How did this affect the mission?

SUBJECT	COMMENT
01	No Comment.
02	No Comment.
03	No Comment.
04	No Comment.
05	No Comment.
06	No Comment.
07	Would enhance mission, assuming operator was trained in FLIR recognition.
08	Nice to have SPOT following HRGM.
09	No Comment.
10	No Comment.
11	No Comment.
12	Increased uncertainty of target #2 (without chance to look at 3 & 4).
13	No Comment.
14	No Comment.
15	No comment.

11. IN-THE-LOOP-CONTROL: How did this affect your "feeling" of in-the-loop control?

SUBJECT	COMMENT
01	No Comment.
02	No Comment.
03	See comment on run 3. : If I can still look at possible targets that the ATC does not identify. (If I can only look at ATC target circles, and can't call up the 2.5' on other targets, then the MIL is degraded.)
04	No Comment.
05	No Comment.
06	No Comment.
07	Slightly enhanced in that you can more readily recognize the truck.
08	No Comment.
09	No Comment.
10	No Comment.
11	No Comment.
12	Definitely need ability to select target areas to ensure I'm not selecting wrong target.
13	Too much is not in my control. Register error between radar SPOTS reduced confidence FLIR WCS showing designated area.
14	No Comment.
15	No comment.

SENSOR AVIONICS ALTERNATIVE

1. If you were using real avionics in an actual aircraft, how well would you be able to DETECT (target position)?

SUBJECT	COMMENT
01	No Comment.
02	No Comment.
03	No Comment.
04	No Comment.
05	No Comment.
06	No Comment.
07	Wouldn't use FLIR to detect, seems to narrow in image size.
08	Much better detection with FLIR.
09	No Comment.
10	No Comment.
11	No Comment.
12	No Comment.
13	Two sensors would be better than one.
14	No Comment.
15	No comment.

2. If you were using real avionics in an actual aircraft, how well would you be able to LOCATE (handoff target position/location to a weapon)?

SUBJECT	COMMENT
01	No Comment.
02	No Comment.
03	No Comment.
04	No Comment.
05	No Comment.
06	No Comment.
07	Same as above.
08	FLIR imagery can help locate target, being able to pass off the location is dependent upon the accuracy's of the FLIR vs SPOT and how the two are integrated.
09	No Comment.
10	No Comment.
11	No Comment.
12	No Comment.
13	Hard to do manual search with extra sensor.
14	No Comment.
15	No Comment.

3. If you were using real avionics in an actual aircraft, how well would you be able to IDENTIFY (target)?

SUBJECT	COMMENT
01	No Comment.
02	No Comment.
03	2 different ID methods -- great!
04	No Comment.
05	Could move FLIR to survey area.
06	No Comment.
07	Excellent capability to ID once target is located.
08	Better capability than with SPOT alone.
09	No Comment.
10	No Comment.
11	No Comment.
12	No Comment.
13	If you get target on both devices, it would be a compelling ID!
14	No Comment.
15	(Added at end) If I had control of the FLIR, I think I could have much more confidence in "IDENTIFICATION" of targets. It would increase SA with a bit of slow-down in performance of other tasks.

4. WORKLOAD: How might this avionics capability affect your workload?

SUBJECT	COMMENT
01	No Comment.
02	May affect decision making which sensor to use to make final crosshair placement.
03	Not as good as with ATC. About even with Baseline. (less uncertainty, but more to do).
04	No Comment.
05	More to look at to increase confidence or targeting decision.
06	Need to manually point the FLIR would increase overall workload.
07	Decrease because it would aid in operator correctly identifying target.
08	Steering the FLIR to the area won't be much more work - the sensor change mentally is different.
09	No Comment.
10	No Comment.
11	No Comment.
12	No Comment.
13	Hard to manage sensor without manager assistance.
14	No Comment.
15	No Comment.

5. TRAINING: What effect would this avionics capability have on training requirements?

SUBJECT	COMMENT
01	No Comment.
02	No Comment.
03	No Comment.
04	No Comment.
05	No Comment.
06	New control/system.
07	Again FLIR training would be essential.
08	No Comment.
09	No Comment.
10	No Comment.
11	No Comment.
12	No Comment.
13	Tough to learn.
14	No Comment.
15	No Comment.

6. CONFIDENCE: How might this affect your mission confidence?

SUBJECT	COMMENT
01	No Comment.
02	See 4 above, Could go either way which is best calibration. More sensors the more complicated.
03	No Comment.
04	No Comment.
05	See #4 answer: More to look at to increase confidence or targeting decision.
06	No Comment.
07	Consistent target ID with FLIR would greatly increase confidence of system.
08	No Comment.
09	No Comment.
10	No Comment.
11	No Comment.
12	No Comment.
13	Confirmation in marginal situations.
14	No Comment.
15	(Added at end) If I had control of the FLIR, I think I could have much more confidence in "IDENTIFICATION" of targets. It would increase SA with a bit of slow-down in performance of other tasks.

7. SITUATIONAL AWARENESS: How might this affect your SA?

SUBJECT	COMMENT
01	No Comment.
02	No Comment.
03	No Comment.
04	No Comment.
05	No Comment.
06	No Comment.
07	No Comment.
08	Need to be more positionally aware of target environment when steering FLIR.
09	No Comment.
10	No Comment.
11	No Comment.
12	No Comment.
13	More info the better.
14	No Comment.
15	(Added at end) If I had control of the FLIR, I think I could have much more confidence in "IDENTIFICATION" of targets. It would increase SA with a bit of slow-down in performance of other tasks.

8. PACING: How might this affect your mission pacing?

SUBJECT	COMMENT
01	No Comment.
02	Too much at times.
03	No Comment.
04	No Comment.
05	No Comment.
06	No Comment.
07	Would probably allow more time for other tasks.
08	Marginally slower.
09	No Comment.
10	No Comment.
11	No Comment.
12	No Comment.
13	This would be a very busy system.
14	No Comment.
15	No Comment.

9. FALSE ALARMS: How might this affect false alarm occurrence?

SUBJECT	COMMENT
01	No Comment.
02	No Comment.
03	No Comment.
04	No Comment.
05	No Comment.
06	No Comment.
07	Prevent some false ID target selection.
08	No Comment.
09	Reduce occurrence.
10	No Comment.
11	No Comment.
12	No Comment.
13	Added good info would reduce FA.
14	No Comment.
15	No Comment.

10. SENSOR MANAGER/ATC: How might this affect the mission?

SUBJECT	COMMENT
01	No Comment.
02	No Comment.
03	No Comment.
04	No Comment.
05	No Comment.
06	No Comment.
07	Same as above.
08	No Comment.
09	No Comment.
10	No Comment.
11	No Comment.
12	Would give me higher confidence with target detection without always having to have the FLIR (which can actually cause more uncertainty).
13	No Comment.
14	No Comment.
15	No Comment.

11. IN-THE-LOOP-CONTROL: How might this affect your "feeling" of in-the-loop control?

SUBJECT	COMMENT
01	No Comment.
02	No Comment.
03	No Comment.
04	No Comment.
05	No Comment.
06	No Comment.
07	Would let user feel a better decision regarding target ID could be made, i.e. "empowering" the user with critical information (target image) to make decisions.
08	No Comment.
09	No Comment.
10	No Comment.
11	No Comment.
12	Let's me decide if I want/need the FLIR for more information. Too much information in Run 4 increased the stress.
13	No Comment.
14	No Comment.
15	No Comment.

OVERALL IMPRESSIONS

1. REALISM: How realistic were the mini-mission scenarios?

SUBJECT	COMMENT
01	You have a lot more time since there are no other aircraft items to monitor.
02	Too laid back. Compress time or deviate for some of the depicted threats.
03	Don't really have anything to base off of.
04	No Comment.
05	No Comment.
06	No Comment.
07	Realism was OK, could have been more realistic with background noise (engines, radio traffic, threat warning systems, etc.).
08	NAV does not end when targeting starts. Navigating toward the target area while scanning the targets may be different.
09	No Comment.
10	No Comment.
11	No Comment.
12	No Comment.
13	Needs to be part of full-task. Not enough other things on my mind.
14	No Comment.
15	No Comment.

2. PREPARATION TIME: Was adequate time given for mission preparation ("ground school")?

SUBJECT	COMMENT
01	No Comment.
02	Excellent preparation! Great pre brief.
03	Not rushed at all.
04	No Comment.
05	No Comment.
06	No Comment.
07	Explanations were adequate. As always, you learn more by doing, so learning curve was steep.
08	No Comment.
09	No Comment.
10	No Comment.
11	No Comment.
12	Have testee talk tester though a normal sequence of events before getting into sim ("chair fly") - especially NAV update because of extra button to push.
13	No Comment.
14	No Comment.
15	I felt I understood the scenario.

3. EXPLANATION: Was adequate explanation given for mission tasking?

SUBJECT	COMMENT
01	No Comment.
02	No Comment.
03	Nice, step-by-step explanations. Previous familiarity with terms and concepts did help.
04	No Comment.
05	No Comment.
06	No Comment.
07	No Comment.
08	No Comment.
09	No Comment.
10	No Comment.
11	Could have used a better explanation of workload as it pertained to that crew station (i.e. what to evaluate against).
12	No Comment.
13	No Comment.
14	No Comment.
15	I knew what to do and how.

AIR CREW BACKGROUND AND EXPERIENCE QUESTIONNAIRE

SUBJECT	SEX	AGE	AIRCRAFT TYPE	HSD EXPERIENCE	SENSOR EXPERIENCE
01	M	24	T-37; T-38	No	No
02	M	45	B-52G/H; B-1B	No	Yes- SAR and FLIR
03	F	26	T-37; T-38; C-18	No	No
04	M	33	T-37; T-38; B-52G/H	Yes	Yes-EO:EVS(STV) and FLIR
05	M	43	T-29; B52-G/H	No	Yes-Real beam ASQ-38; EO:EVS(STV); FLIR
06	M	41	B-52	Yes	Yes-EO:EVS(STV) and FLIR
07	M	25	T-38	No	Yes-Real beam T-42A
08	F	23	T-37; T-38; C-5	Yes	Yes-Ground based radar and EO; FLIR
09	F	24	T-38	No - Yes, if Horizontal situation display is the same as an HSI (instrument) in cockpit	No
10	M	26	T-38	No	No
11	M	26	T-38	No	No
12	M	27	T-38	No	No
13	M	45	B-52	Yes	Yes-real beam B-52; SAR (simulations); ERS/B-1B; EO:EVS(STV); B-52; IR:FLIR, MH53J
14	M	30	T-37	No	Yes-Weapon-mount EO:AGM 65 A/B; Weapon-mounted IR:AGM65D; Pave Penny; AIM9L
15	M	28	T-38; T-37; Cessna	No; only HSI	No

U. S. Government Printing Office 1995 650-075/00126